

Vacuum at the heart of Europe

A sudden withdrawal of support by the European Union for key scientific institutes could be a major setback for the continent's biological research and is yet another example of a lack of scientific vision at the European level.

Stick a pin anywhere on a map of Europe, and the country's head of state will at some point have sung fine words about how biology and biotechnology are top research priorities that, along with information technology, will fuel the industrial and medical revolutions of the next century. Such speeches inevitably include a warning that Europe needs to be "competitive with the United States and Japan" and that cooperation is one way of achieving this. Regrettably, a Scientifically United (or even Coherent) States of Europe is still a long way off, as illustrated by the disarray behind an impending crisis at the European Bioinformatics Institute (EBI) in Cambridge and other European-funded life-science centres in Europe (see pages 3 and 12).

EBI was launched as a joint venture in 1994 by the European Union (EU) and the European Molecular Biology Laboratory. But EU member states have now told the European Commission that it should no longer fund research infrastructure. Unless something happens soon, EBI, already a key facility for European molecular biologists, will wake up to the year 2000 minus half of its already pitiful budget of \$8 million — its US counterpart, the National Center for Biotechnology Information, will spend \$19 million this year.

Yet the dire consequences of this policy were not even discussed in the committees of representatives from the member states that devised it. If this Kafkaesque affair has any merit, it is that it has exposed the absence of a clear mechanism for the planning and support of research infrastructure at the European level. The European Commission has a clear mandate to coordinate what is better done at this rather than at the national level. But running large infrastructure on the short-term project grants the commission hands out is no substitute for long-term stable funding commitments.

In practice, such efforts have been driven forward largely through multilateral coordinated actions by the member states themselves. In the past, such deals have created pan-European institutions, such as the European Laboratory for Particle Physics (CERN), the European Synchrotron Radiation Facility and the science programme of the European Space Agency. These show that, when Europe gets its act together, it can excel.

But many of these programmes were launched by a generation of more visionary 'European' decision-makers. That vision currently seems to be lacking. The problem on infrastructure, as the EBI fiasco vividly illustrates, is that the member states have no generalized mechanism for making multilateral policy on scientific issues outside the commission. The patchy and ill-coordinated development of future European neutron and light sources illustrates the point. A coherent approach would have been for Europe to have first assessed its needs as a whole, and to have then worked out how best and most cost-effectively it could meet them collectively.

What is needed in Europe is not a new bureaucracy to deal with research infrastructure, but for science ministers and funding agencies to start thinking in terms of European solutions, and acting accordingly. At present, cooperation only takes place in earnest when the sums involved make this unavoidable, and even then it is often the lowest-common-denominator form of a project that prevails. Despite its enormous scientific importance, it is perhaps the relatively low costs of EBI which explain its low political profile so far.

The need for European cooperation is greatest in areas where the costs alone do not appear to make it essential. Most of Europe's research funds are scattered across national research councils and other funding agencies. Although local funding is often appropriate for investigator-driven research, the duplication of effort is frequently excessive, while many disciplines could benefit from greater concentration of much larger amounts of money.

But focusing substantial biomedical funds on strategic priorities inside a European National Institutes of Health, for example, is not for tomorrow. The EBI crisis should prompt the science agencies and their paymasters to complement their fine words about making European life sciences competitive on the world stage with vision, leadership and action. A decision by member states at the forthcoming meeting of the European Molecular Biology Laboratory to provide full funding for EBI as a short-term measure — and to register a firm commitment to negotiate a mechanism for the long-term support of this and similar initiatives — would be a good place to start. ■

Fiction's futures

A forward-looking series of articles depends at least as much on imagination as reality.

Nothing dates as quickly as a prediction. That is why we have asked authors of our new weekly 'Futures' series, briefed to think about the scientific advances of the millennium to come, to concentrate on enjoying themselves rather than presenting anything especially serious or accurate. After all, when *Nature* was first published, 130 years ago today, predictions of the twenty-first century might have portrayed women in crinolines riding around in steam-driven airships.

This pursuit of happiness explains why we have asked, in the main, science-fiction writers to share their thoughts. As a genre, science fiction has purposes besides entertainment. It provides a medium in which writers can express not their predictions but their preoccupations with the present day. Moreover, such professional writers are

perhaps better equipped than scientists to understand and convey how technological changes will affect the way we live, in all sorts of ways besides the mechanical. Nevertheless, we do expect scientists to contribute to this series, and we shall also be publishing other researchers' anticipations in a supplement at the beginning of next month.

In the weeks to come, the Futures series will present new writing from some of science fiction's household names — and some excellent new writing from authors of whom you might not have heard. For the first exercise in this new venture, *Nature* is delighted to publish an original contribution from Sir Arthur C. Clarke, perhaps science fiction's most famous name of all. You can find it on page 19. ■





Splitting up
Stanford/UCSF
hospitals merger fails
to convince critics
p5



Relativity values
NASA gravity
experiment hits
technical snag
p7



Extreme reaction
US primate
researchers facing
protests and threats
p7



Fellow feeling
'Marie Curie'
fellows seek a
political profile
p9

Life science facilities in crisis as Brussels switches off funding

Paris

Several major European life-science research facilities face financial crisis following the European Commission's decision not to fund their operational costs out of the fifth five-year Framework programme (FP5).

One facility directly affected is the European Bioinformatics Institute (EBI) in Cambridge, which forms Europe's main database infrastructure for molecular biology and biotechnology. An outstation of the European Molecular Biology Laboratory (EMBL), EBI has relied on the European Union for almost half its budget, but now finds that this tap has suddenly been switched off.

Other facilities affected are the recently created European Mouse Mutant Archive, based at Monterotondo near Rome (see next page), and the *Drosophila* stock centres at Umeå in Sweden and Szeged in Hungary.

Under instruction from member states, the European Commission published a new rule in March stating that core funding and operational costs for infrastructure should not qualify for support from FP5, which runs from 1998 to 2002. Member states wanted to control investment decisions on large science facilities themselves, and preferred such funding to be provided directly.

But EBI failed to recognize the rule's significance, especially since research infrastructure was made a priority in FP5, with funding rising from 200 million euros (US\$212 million) in the previous Framework programme to 500 million euros for this one. EBI had not read the small print specifying that such funds would be used only for research projects and to fund researchers from one country working at facilities in another.

Realization dawned over the summer, when the European Commission rejected EBI's applications for Framework funds. No one doubts the scientific excellence of EBI, says one commission official. "They were simply outside the scope of funding."

Embarrassed commission officials are now advising EBI on how to rewrite their proposals to include areas that will meet the new funding criteria. But one official admits that sleights of hand, such as dressing up



Hard times ahead? The European Bioinformatics Institute in Cambridge faces a funding problem.

infrastructure costs as research projects, are no longer an option. A spate of financial scandals prompted the commission to resign *en masse* earlier this year, and in the new squeaky-clean climate the commission is applying the rules to the letter.

As a result, funding for EBI from existing European Union contracts will dry up at the end of the year, and EBI accountants estimate that the shortfall will form 44 per cent of its budget. If the immediate cash-flow problem is not resolved, "we will have to abandon major projects like the DNA database, the draft human genome, the macromolecular structure database and the

microarray expression database," warns Graham Cameron, who jointly heads the institute with Michael Ashburner, professor of biology at the University of Cambridge.

The crisis has come at a bad time for EBI. Demand for its services is growing at 15 per cent every month, while the *Drosophila* genome is likely to be available in February, with a draft of the human genome in the spring (see *Nature* **401**, 729–730; 1999).

The need for EBI to expand is widely supported (see Correspondence, page 12). The institute and many in the community it serves believe its annual budget should be doubled from its current \$8 million if it is to remain competitive with its US equivalent, the National Center for Biotechnology Information, which has a budget of \$19 million.

The EMBL board will meet later this month to try to drum up money from its member states to tide EBI over until a permanent solution is found. Most observers are optimistic that the money will be found, as they argue that it would be unthinkable to let EBI sink.

"Everyone realizes that this is really a small amount of money for so many countries to put together to keep Europe competitive," says Julio Celis, chairman of the EMBL

South Africa says AIDS drug 'toxic'

Cape Town

South Africa's president Thabo Mbeki last week said the country would not take the "irresponsible" step of supplying AZT (zidovudine) to HIV and AIDS sufferers until the drug's safety was established.

The statement is being seen as an attempt to justify the government's tardiness in making AZT available in state hospitals, even to rape victims and pregnant women.

Mbeki claimed that legal cases were pending in South Africa, the United States and Britain against the use of AZT on the grounds that it was harmful.

But this has been strongly denied by Peter Moore, medical director for sub-Saharan Africa for Glaxo Wellcome, the

drug's suppliers, who have requested a meeting with Mbeki to clarify the issue. The company has been negotiating with the government for the past three years over the price of supplying the drug to state hospitals.

Mbeki's statement was made in his first address to the National Council of Provinces since he became president in June. He argued that a large body of scientific literature claimed that AZT is so toxic as to be a health hazard. He had asked national health minister Manto Tshabala-Msimang to investigate this, but said that, until her investigation was complete, it would be irresponsible for the government to ignore researchers' warnings.

Michael Cherry

council and head of the Danish Centre for Human Genome Research in Århus.

But one official at the European Molecular Biology Organization (EMBO) is more cautious: "Given the vagaries of European politics, I would not be overly optimistic that the member states will say, 'fine, let's pay up.'"

Since EBI was created in 1994, it has been funded five parts by EMBL, five parts by the European Commission and two parts by industry. But there is no alternative plan now that the commission has pulled out.

Fotis Kafatos, director of EMBO, says that the fundamental issue is whether Europe will invest in essential structures for molecular biology and biotechnology, and that the challenge for European science is to co-ordinate national, multilateral and pan-European systems.



Kafatos: calls for more coordination.

"There is no clear mechanism for creating pan-European infrastructures that are globally competitive, because the funding is so fragmented," adds Cameron.

Commission officials say that, in principle, the treaties regulating the European

Union give it a clear mandate to plan and fund research infrastructure — the principle of subsidiarity says it should support activities that are best carried out at the European level rather than at the level of member states.

But the political reality is that member states are loath to give the commission such powers. The favoured model is one in which transnational initiatives are organized under multilateral accords between member states — the so-called 'variable geometry'.

According to one commission official, member states have been keen to keep decisions on big science facilities, such as the European Spallation Source and new synchrotrons, out of the commission's hands. Member states believe they are better at controlling budgets and accountability.

The challenge is to devise a mechanism for stable long-term funding to secure EBI's future as a globally competitive organization, says Ashburner. The United Kingdom is said to be ready to lead efforts to achieve this.

Britain is unlikely to contest the ruling against supporting EBI, as it was among the countries most against funding infrastructure directly, says a commission official.

Discussions are taking place on several levels. Britain's Medical Research Council is brokering discussions with other European

research councils for a rescue plan, and is seeking support from governments.

The Wellcome Trust, which contributed half of the initial construction costs of EBI (see *Nature* **361**, 383; 1993), is also being mentioned as a possible backer, although its support may depend on governments agreeing a long-term plan for EBI.

A meeting of the EMBL council in March will seek support for EBI from its member states. EMBL's current five-year plan for its programme and budget ends in December 2000 and, if a solution is to be found within EMBL, European member states will need to agree to a funding envelope for EBI in the next five-year plan.

"We need a long-term political commitment," says Ashburner. One option member states may consider is to fund EBI as a separate entity within a new multilateral accord.

Celis believes that the crisis has precipitated a long-overdue discussion on securing a stable and adequately funded structure for EBI and for other European-level facilities. Project funding is no way to run large infrastructure projects, he says. "What is needed is a mechanism involving stable partners, stable financing and long-term commitments."

Declan Butler

Framework funding ban hits mouse mutant archive

Munich

The financial crisis facing the European Bioinformatics Institute (see above) is also threatening the European Mouse Mutant Archive (EMMA). The archive, which has its main site at Monterotondo near Rome, has had its application for European Union (EU) infrastructure funding turned down.

The application was coordinated by Davor Solter, a director of the Max Planck Institute for Immunobiology in Freiburg, and a member of the programme committee that selects strains to be held at EMMA and its nodes in France, Britain and Sweden.

The funding request covered "basic activities required for acceptance of mutant strains, freezing of sperm and embryos, and molecular characterization of the strains".

EMMA was set up with grants totalling 3.9 million euros (US\$4.1 million) from the biotechnology infrastructure programme within the EU's fourth Framework programme (FP4), and with support from national organizations. It was created in response to the demand for mutant mice for the study of the function of sequenced genes.

Scientists and European Commission officials involved in negotiating EU grants in 1996 and 1997 expected continued funding in subsequent EU framework programmes.

But even then, EMMA scientists viewed the short-term nature of the funding as

unsatisfactory for a facility that cannot be temporarily closed pending new grants.

"You can't put the mice on a shelf and ask them to wait for the next grant before they get their next feed," says Solter.

Now that EU funding has dried up, this dissatisfaction has turned to panic. The FP4 funds ran out in spring this year, at around the time that EMMA officially opened (see *Nature* **398**, 183; 1999). The archive is being temporarily funded by the Italian research council, the CNR, although it was nervous about funding building and running costs.

Solter did not realize that the wording of the programme on which he based his application, which excludes "support for the construction and operation of research infrastructures, as well as for the collection of data", meant that running costs were excluded. "I sort of assumed 'operation' was referring to water and electricity," he says.

Commission officials are encouraging EMMA, like EBI, to reapply for FP5 money for research projects, which they say is all that the infrastructure programme can cover. There is certainly plenty of money available: funding for infrastructure within the life-sciences programme of FP5 totals 60 million euros, of which 12 million euros was available for the first round, but only 5 million euros has been allocated.

But a more stable solution is also being



sought. "We need a constructive and timely alternative, such as a direct contribution from the European Union to operate EMMA," says Glauco Tocchini-Valentini, director of the CNR Institute of Cell Biology at Monterotondo, and CNR representative on EMMA's science-policy committee.

One commission official thinks that EMMA may be harder to rescue than EBI, which is well established and has multi-lateral support within EMBL. In contrast, he says, EMMA is covered by one national research council, and it may need member states to rally to its cause and create a new multilateral legal entity.

Alison Abbott

Californian universities scrap hospitals merger

San Diego

Stanford University and the University of California at San Francisco (UCSF) announced last week that they are to dissolve the merger of their hospitals. This ends an experiment that some academic staff felt was threatening research and teaching at the two institutions.

The move for the divorce came last week when Stanford president Gerhard Casper notified University of California officials that he had "with great anguish" concluded that the attempted merger of the medical staffs was doomed. "Energy has been sapped and great weariness has set in," said Casper in a letter to University of California president Richard Atkinson, who was out of the country. "Many fear paralysis."

Commenting on Stanford's unexpected withdrawal, Michael Bishop, chancellor of UCSF, said his university had been prepared to try to continue to make the troubled health system work. But, given Stanford's decision, UCSF will now engage in "a cordial dissolution process". The University of California Board of Regents must still approve the dismantling of the two-year merger, but little opposition is expected.

With an annual operating budget of \$1.5 billion, and extensive medical and research expertise at four hospitals in Palo Alto and San Francisco, the joint venture had been intended to create a single enterprise with great financial and intellectual strength. It was hoped it could capitalize on discoveries by bringing them rapidly from the bench to the bedside, thereby bolstering revenues. There were also hopes that the combined medical expertise would draw more patients.

But, in practice, the effort to merge the hospitals ran into many problems. Medical staff had balked at combining services, reportedly because of concerns about loss of power, individual university identity and pay. And the merger was also hit by reduced government reimbursement for health care which contributed substantially towards the \$86 million losses during the past fiscal year.

The merger has long been controversial among staff at both institutions, largely because some faculty members felt they had little voice in decision-making. In the weeks before Stanford's move, two faculty polls at UCSF — one by the faculty association and the other by the academic senate — each found a small majority in favour of dissolving the merger, by 52 and 53 per cent, respectively.

Warren Gold, a physician who chairs the UCSF faculty association, described the polls as "evidence the university leadership is out of touch with the faculty". Gold, who has



Merger doomed: Stanford (above) has pulled out of plan to run its hospitals with those of UCSF.

opposed the merger since its inception because he felt its goals were unrealistic, says his association poll of 535 of UCSF's 1,300 faculty members found that 24 per cent felt their ability to conduct research had "worsened" since the merger, while 26 per cent said their ability to teach had also been harmed.

However, Bishop says the senate's poll revealed that 64 per cent of faculty members said they were in favour of the merger provided it was handled properly. Lawrence Pitts, a neurosurgeon who chairs the UCSF academic senate, says the "faculty was disenchanted with the extraordinarily poor administration" of the combined hospital enterprise.

Keith Yamamoto, chairman of the department of cellular and molecular pharmacology, says that, although most basic researchers at UCSF were not affected by the contentious merger environment, physicians with doctorates — particularly young researchers — felt their scientific efforts were threatened by the pressure to produce revenue.

Although there was no poll at Stanford, faculty leaders reported concerns last summer that the requirement to focus excessively on revenues from clinical work could have a negative impact on research.

Casper acknowledges that university officials had always been aware that there was "no overwhelming support of the faculty" for the merger, but had proceeded out of a desire to create a stronger institution.

Overseeing the dissolution will be one of Casper's last major challenges, as he will be stepping down as president next August. He estimates that it will take three to four months to separate the hospital enterprises. The cost will be substantial.

Rex Dalton

Silicon pioneer funds biomedical engineering centre

San Diego

James H. Clark, the founder of Silicon Graphics Inc. and the Netscape Communications Corp., announced last week that he is donating \$150 million to Stanford University for a new biomedical engineering facility.

According to university officials, the grant to Stanford, where Clark was a professor between 1979 and 1982, is the largest gift to the university since it was founded and is one of the biggest bequests ever to a US university.

The Clark Center for Biomedical Engineering and Sciences will be built on campus near the university's science facilities, with completion expected in 2002. Known on the Stanford campus as 'Bio-X', the facility has been in the planning stages for nearly two years, beginning as a grassroots effort by faculty members keen to create a focal point for multidisciplinary research.

It will house about 400 scientists and technicians. Some 50 faculty members, about 30 per cent of whom will be new recruits, will conduct research in bioengineering, biocomputing, neuroscience and imaging at the molecular, cellular and system level.

The complex is to include a 'reality centre', in which displays on the ceiling, floor and walls will allow the projection of images to show visitors the inside of structures such as cells or arteries.

"I chose to do this because of my academic roots, and Stanford is a great place," says Clark. "If you're allowed to be in an academic setting and create the springboard of a business effectively without undue impediments, then you have an obligation to respond in kind."

John Hennessy, Stanford's provost and a former colleague of Clark in the Department of Electrical Engineering, says discoveries in genetics and cellular biology, with advances in computing and the miniaturization of devices, "will provide incredible opportunities for advances in biomedicine, bioengineering and bioscience".

Neuroinformatics is of particular interest to Clark, who decided that the most effective way to contribute was to fund a broader effort, with biology as the central focus. During his career, Clark moved from engineering to physics and then to computer science. His first major success was a computer chip, called a geometry engine, that was the foundation of Silicon Graphics.

R. D.

NIH panel to limit secrecy on gene therapy

Washington

The US government advisory committee that monitors experiments in human gene therapy plans to act early next month to stiffen its guidelines on the reporting of deaths and adverse events.

The action by the Recombinant DNA Advisory Committee (RAC) of the National Institutes of Health (NIH) follows attempts by a leading researcher and a drug company to keep confidential a death and serious side effects, respectively, suffered by gene therapy patients.

Current RAC guidelines require scientists to report any serious adverse event immediately to the NIH's Office of Recombinant DNA Activities (ORDA), which staffs the RAC. But they do not say whether such reports can be marked 'confidential', and thus kept off the RAC's public agenda. The committee aims to remove this option.

"Trying to hold these confidential may be of short-term benefit for some reasons for the company. But in the long term it's [not] in the best interests of the field," says Claudia Mickelson, the RAC chair and institutional biosafety officer at the Massachusetts Institute of Technology.

The RAC decision followed a meeting in September, at which it was presented with a letter sent to ORDA in May 1998. In the letter, stamped confidential, Ronald Crystal, a gene therapist at the New York Hospital-Cornell Medical Center, reports — but gives no details of — the death of a 61-year-old man during a gene therapy trial aimed at inducing the growth of new heart blood vessels.

Crystal states that the death, 40 days after



Keeping it public: a patient awaits gene therapy for muscular dystrophy.

gene therapy was administered, was not linked to the therapy. He requests that the letter "be kept confidential and not part of the public record". Crystal reported fully on the case at the 1998 conference of the American Heart Association, and in the 3 August 1999 issue of the journal *Circulation*.

Crystal says that he asked for secrecy to protect the patient's family. He adds that the RAC is understaffed, and ill-equipped to judge whether the death was related to the gene therapy. "I was very concerned about releasing confidential patient information in the context of a group that has neither the appropriate secrecy nor the expertise and manpower to evaluate it."

Crystal wrote the letter around the time that the Rockville-based biotechnology company that he founded, GenVec, filed to make an initial public stock offering. But he

says that the two events were unrelated, and that his interest in the company was 1.5 per cent of an enterprise valued at the time at about \$70 million. Company officials ultimately decided not to go public.

In a separate incident, the drug company Schering-Plough reported to ORDA in September, describing two patients who had developed side effects in gene therapy trials for colorectal and ovarian cancer. "Some of this information is proprietary competitive information," Bob Consalvo, a company spokesman, said in a statement last week.

NIH officials say that these are not isolated cases. In a research area once funded largely by public money, but now almost totally reliant on industry, a number of investigators and their sponsors are asking for elements of protocols — and sometimes entire protocols — to be confidential.

The RAC has always allowed proprietary information to be kept confidential. But secrecy requests involving deaths and side effects go against the spirit of a fledgling field that has sought to gain public trust through openness. In September, for example, the death of an 18-year-old man in a gene therapy protocol was fully reported by James Wilson, of the University of Pennsylvania (see *Nature* **401**, 517; 1999).

NIH officials, however, tread a fine line in seeking to keep the field transparent. If too much is revealed, they risk scaring away commercial sponsors. "The challenge is to figure out a way that we can put patients first and still preserve the commercial interest in this field that has vitalized it," says Amy Patterson, the ORDA director. **Meredith Wadman**

Protests grow over interrogation of Ukraine scientists

London

Ukraine's secret services have widened their investigation into staff at the Institute of Biology of Southern Seas (IBSS), following the arrest of marine biologist Sergei Piontkovski on charges of sending 'secret' information abroad and handling foreign currency illegally. Meanwhile a growing number of international organizations have protested about the action (see *Nature* **401**, 835; 1999).

Ten IBSS staff have now been questioned by Sebastopol security services (USB). All were participants in either a UK-funded Darwin Initiative project or a European-funded INTAS project. The grants from these projects, for the collection of biodiversity data, appear to be the focus of the intelligence operation. INTAS says other grant recipients in southern Ukraine have been asked to report to USB.

INTAS, which promotes cooperation with scientists from the independent states of the former Soviet Union, said last week that it was "shocked" at Piontkovski's detention. The Ukrainian Ministry of Science and INTAS have confirmed that Piontkovski's work was done in agreement with the Ukrainian government and in accordance with the law. An official reaction from the government is still awaited.

Elisa Muñoz, of the committee on scientific freedom and responsibility of the American Association for the Advancement of Science, says she will ask the committee's network of scientists to write to the Ukrainian authorities to request Piontkovski's unconditional release.

The European Commission has requested an explanation from the same authorities, and the International Council for Science's committee on oceanic research



Piontkovski: fears a charge of espionage.

is writing to the Ukrainian Academy of Sciences.

A friend of Piontkovski's has set up a website on the case (http://www.geocities.com/sep_case/). On it, a fax from the Ukrainian science ministry to the head of the security services warns that the action could have "negative consequences for the international scientific collaboration of Ukraine".

Piontkovski was held in USBU offices for four days and subjected to interrogations that he says were "long term and with unpleasant dialogue with a lot of threats that [the charge] would fall into the category of espionage actions". **Natasha Loder**

NASA relativity probe facing costly delay

Washington

A controversial project by the US space agency NASA to test Einstein's general theory of relativity in Earth orbit has run into a technical snag that is likely to increase costs and delay its launch. Some \$21 million or more could be added to the \$535 million price tag, and the launch — originally planned for next October — could be put back by several months at least.

Project managers at Stanford University in California believe that thermal problems with the Gravity Probe B (GPB) spacecraft should be relatively easy to fix, but the uncertainty is worrying NASA science officials. "We're colouring it red," says Alan Bunner, who heads the agency's programme on the structure and evolution of the Universe.

GPB, also known as the Relativity Mission, will try to measure precisely the 'frame dragging' effect, whereby a massive rotating body such as the Earth is thought to drag space and time along with it as it turns. The effect is expected to show up as a slight deviation in the alignment of four exquisitely sensitive gyroscopes spinning inside a super-cooled helium dewar.

During recent ground tests, however, engineers discovered that heat is not being conducted out from the interior of the spacecraft as it should. If the problem were left unfixed, cryogenic helium in the thermos-like dewar would boil off in orbit within a few months, ending the 16-month mission prematurely.

The culprit appears to be an epoxy resin that has come unbonded from a titanium liner inside the probe, severing an important pathway for conducting heat. Stanford engineering professor Brad Parkinson, one of three co-investigators for GPB, says "the probability is 98 per cent that this is the problem," and that "we can easily see a solution" — adding copper pins and epoxy to restore the thermal pathway.

But this requires removing the probe from the dewar. This would delay the launch from next October to at least May 2001, and probably later if an additional margin is built into the schedule as the Stanford team would like. NASA expects to decide the new launch date this month.

First conceived at Stanford University in 1959, GPB has already had a long and turbulent history. NASA commissioned the first mission study in the mid-1960s, and began funding substantive work on the project in the mid-1980s.

Often criticized as a technologically risky experiment with little support from the wider scientific community, GPB has been threatened with cancellation several times. The most recent hurdle was a make-or-break 1995 recommendation to NASA by a



Getting warm: the dewar housing the gyroscopes may lose its cryogenic helium too quickly.

National Research Council committee that the project should continue.

But the committee was deeply divided as to GPB's merit compared with other space projects. And it cautioned that a surprising result — Einstein shown to be wrong — would probably not be accepted without a repeat mission. NASA has no plans to fly a follow-up mission.

Parkinson says GPB is a target for criticism because fundamental physics "doesn't appear to have a major constituency" at NASA. And he defends the project's four- to six-per-cent budget overrun as not unusual, especially given the advanced technology that had to be developed. This includes extremely precise pointing and manoeuvring and the use of a 'drag-free' orbit.

But the project has overspent its budget by \$20 million in the past two years, says Bunner, which has raised the level of concern at NASA headquarters. An earlier launch delay from March to October 2000 prompted the agency to add oversight staff at Stanford. Parkinson counters that the \$20 million had actually been taken from programme reserves in previous budget years and was never restored.

And he says the new launch delay, if approved, will allow other minor problems to be fixed, including one that would have made one of the four gyroscopes unusable.

At another time, a \$21 million overrun might not have made a dent in NASA's budget. But the problem comes at a bad time for the agency's science office, which is struggling to contain costs in other areas, including its Mars exploration programme.

Ed Weiler, head of NASA's space science, decided last summer to cancel the Challengeron comet mission, in part because of budget pressures caused by expensive delays with the Chandra X-ray observatory (see *Nature* 400, 99; 1999). The Congress also recently earmarked more than \$70 million worth of pet projects to be taken out of the general science budget in the coming year.

Congress also gave Weiler a \$10 million increase in the 'fundamental physics' account. Instead of funding new missions, the money is likely to cover the latest delay with GPB, the longest-running development project in NASA's history. **Tony Reichhardt**

Threats to US primate researchers

Boston

A radical animal rights group calling itself the Justice Department last week mailed threatening letters, with razor blades dangerously positioned inside the envelopes, to scientists at primate research centres across the United States.

The letters, which were sent from Las Vegas, Nevada, with no return address, told the researchers: "You have until autumn of 2000 to release all your primate captives," and added "if you do not heed our warning, your violence will be turned back upon you".

The group's website includes a list of 83 scientists who had been targeted. Letters have turned up at the universities of Harvard, Emory, Tulane, California, Michigan, Minnesota, Oregon and Wisconsin, but no injuries have been reported.

The universities had already been contacted by the Foundation for Biomedical Research, a non-profit organization in Washington that endorses animal research

and monitors the websites of extremist groups.

"The scientific community is not prepared for these kinds of terrorists acts, so we're doing anything we can to protect them," said the foundation's president, Frankie Trull. "Until recently, there had been a sense of complacency regarding animal rights activists, but this incident should help scientists realize the threat to their personal safety and to their ability to conduct research."

The threats are being taken seriously by the universities. The letters have been submitted to the Federal Bureau of Investigation, which is handling the case, and university officials are calling for extra vigilance and, in some cases, heightened security measures.

Protests at Harvard's primate research centre in Southborough, Massachusetts, over the past two summers had already prompted the installation of additional surveillance equipment.



Burning issue: earlier protests against primate research in Portland, Oregon.

“This is an escalation of what has happened in the past, but fortunately no one was hurt,” said Don Gibbons, a spokesperson for Harvard Medical School, which last week received eight letters from the Justice Department.

Meanwhile, extra security guards have been stationed at the facility. Rocks were thrown at a security guard's car during a routine patrol last Thursday, but officials cannot link the incident to the threatening letters.

The Justice Department has not targeted US researchers before, although it has been active in Britain and Canada. Five years ago, it sent letter bombs to European companies, and two years later sent letters containing razor blades to hunting guides and furriers in Canada.

The best-known extremist group, the Animal Liberation Front, has said it was responsible for fire-bombing four vans belonging to a Rhode Island furrier two weeks ago. The group also claimed responsibility for a break-in at a Western Washington University laboratory two days later. Offices were ransacked and several dozen laboratory animals stolen.

“Acts of violence appear to be on the increase,” says Trull. Along with others who support the use of animals in medical research, she fears that activists are resorting to the same intimidation tactics used by anti-abortion extremists.

Andrew Rowan, senior vice-president of the Humane Society of the United States, says he is “unalterably opposed to these tactics. Threatening violence is contrary to the basic tenets of the animal protection movement, which is aimed at protecting all animals, including humans.”

Trull says that, although catching those who sent the letters is a job for law-enforcement agencies, scientists have a part to play. “The scientific community has to explain to the public why animal research is necessary,” she says. “Very few medical advances made this century did not depend, at least in part, on animal research, and that's not likely to change in the near future.” **Steve Nadis**

French scientists resign from Allègre's advisory council

Paris

Two prominent French scientists have resigned from the National Science Council (CNS), a special advisory panel created by the science minister, arguing that it is consulted on major decisions only after they have already been taken by the ministry.

In a move seen as a direct criticism of the leadership style of science minister Claude Allègre, mathematician Yves Meyer and Nobel-prizewinning physicist Claude Cohen-Tannoudji delivered their letters of resignation to Allègre and the prime minister, Lionel Jospin, on 14 October.

“According to [earlier comments by] the national minister for education, research and technology, researchers should return to work in their labs rather than waste time on ineffective committees,” wrote Meyer. “That is what I propose to do.”

Allègre set up the council in October last year to advise him on large-scale projects and research strategy. The move followed the creation of a fund controlled by the ministry, known as the National Science Funds, to be spent on reforms and major projects. This year, it increased to FF565 million (US\$91 million) for the year 2000 budget (see *Nature* **401**, 313; 1999).

But instead of advising the

minister on important decisions, says Cohen-Tannoudji, the council met only three times, and on each occasion listened to reports from the minister on policies or reforms that he had already instituted. “It was more like a briefing session,” he says.

According to Meyer, the last straw came when the minister was quoted in the newspaper *Le Monde* as saying that the council backed Allègre's decision to go along with British plans to build a new synchrotron — two weeks before the council discussed the subject (see *Nature* **400**, 489; 1999).

The decision was taken in preference to plans to build a French machine that had been 10 years in the making. When the issue of the synchrotron was raised at a CNS meeting on 1 October, it met with strong disap-

proval from most council members, according to those present at the meeting.

Vincent Courtillot, director of research at the ministry and Allègre's main adviser, says many decisions taken by the government have been approved by most council members. “Members are free to express their views in the CNS, as was widely done at the 1 October meeting,” he says. “The debates may end with majority and minority views on advice to the government.” **Heather McCabe**



Cohen-Tannoudji: won't waste his time on 'ineffective' body.

Japan plans ethics guidelines

Tokyo

Japan's Ministry of Health and Welfare announced plans last week to set up the country's first ethical guidelines for the management and use of human genetic information for biomedical research.

As an initial step in this direction, the ministry has created an ad hoc working group to address ethical issues on research using human DNA samples, including the protection of personal genetic data, and to discuss measures for obtaining adequate informed consent from patients giving samples.

The working group, whose members include directors of national medical institutes, academics and lawyers, plans to draw up draft guidelines by January. These would be made public as part of the ministry's plan to gather wide opinion from the general public which it intends to incorporate in the final guidelines to be compiled in March 2000.

The move reflects growing concern

among researchers over the potential misuse of genetic information, which, the ministry says, “could not only violate privacy, but also lead to job and insurance discrimination”.

The launch of projects related to the human genome — such as the planned database of single-nucleotide polymorphisms (SNPs), or sequence variations that could be linked to common diseases such as cancer and mental illness — have raised concerns over the confidentiality of research records.

Researchers involved in the SNP programme, which involves collecting a vast number of DNA samples from medical institutions across Japan, have insisted that samples be anonymous to avoid concerns over linked or identifiable samples.

According to the ministry, methods to ensure the anonymity of DNA samples, such as creating codes to identify them, will be the main priority of discussions to be held by the working group. **Asako Saegusa**

Providing support and a voice for Europe's young researchers

A body set up primarily to promote a fellowship scheme for postdoctoral scientists is seeking to play an active role on the political stage.

London

The first open annual general meeting of the Marie Curie Fellowship Association (MCFA) takes place in Brussels next week. The association is hoping to use this meeting not only to raise awareness of the problems facing young European scientists, but also to establish itself as one of the main partners for academia and industry in dialogue on such issues.

The MCFA is open to scientists who have received mobility research training grants from the European Community. Such grants began in 1958 with no distinct identity, being allocated under many different research, training and development programmes. But in 1996, Edith Cresson, then European commissioner for research, announced that research fellowships of the European Union would be called Marie Curie fellowships.

At the same time, she announced the creation of the MCFA. Since 1958, 7,000 fellowships have been awarded and another 8,000 are expected within the fifth Framework research programme (FP5). Around 1,900 people have chosen to join the MCFA, and it already has national groups in every European country.

Practical difficulties in reaching early recipients mean the MCFA can look at recruiting alumni from around 3,200 individuals funded in FP4. But it hopes that a higher proportion of FP5 fellows will join.

Independence

For an organization that only received its first funding in May 1998, nearly 2,000 members is no mean feat. But the MCFA is ambitious, well organized and well funded. In 1998 it received 450,000 euros (US\$483,000) from the commission, and expects 70 per cent of this in 1999. A condition of commission support is that it must find substantial sponsorship funds in its second year. The MCFA hopes to become independent in the next two to three years.

"We see the Marie Curie Fellowship Association playing a key role in the tracking of fellows throughout their career," said Barry McSweeney, head of the Marie Curie Fellowships Programmes in the European Commission. The association also gives the commission valuable input for its monitoring of programme implementation and for planning future programmes.



Bright future: young researchers hope to follow the pioneering footsteps of Marie Curie (right).

Laure Ledoux, president of the MCFA, says that, although the commission is funding the association, it is "independent to create its own aims". One of the main reasons MCFA fellows join, she adds, is the potential it offers for them to link up with researchers in other fields.

"As a young scientist, it is difficult to assert oneself in an ocean of experts in the various research fields," says Ana Lisa Vetere Arellano, a fellow working for a British engineering consultancy firm. "The fact that I am part of the MCFA has helped me acquire a bird's-eye view of ongoing research of my peers throughout Europe."

Fellows also say they hope to meet more experienced researchers and those likely to play key roles in the future. "As far as we know, we are becoming the biggest young researchers' association in Europe. Now, that might be very interesting for anybody's future," says Ana Cerdano, a fellow based at the Sanger Centre near Cambridge in the United Kingdom.

The MCFA can also provide practical assistance to fellows, from help with the tax system of a new country, to what to do on arriving at a university with a family. Members share information about jobs at their university on a joint mailing list, and the association holds a recruitment database.

The association can provide practical assistance, such as help with the tax system of a new country.

Although the commission was primarily seeking, in creating MCFA, to promote Marie Curie fellowships, members are keen for it to have scientific objectives. "The intellectual capital is being trained by the commission and we felt that this international network could capitalize on this," says Ledoux.

The MCFA intends to achieve this by organizing science-based workshops, joint activities with other organizations, such as the European Science Foundation, and by producing *The Annals of the MCFA*, an interdisciplinary journal publishing highlights of the "outstanding research performed by Marie Curie fellows during their fellowships".

The MCFA is actively developing its links with industry. In February, Unilever will sponsor an MCFA meeting on innovation in the biosciences, and thinks it will be a good opportunity to build links with young academics, to find interesting new work, and to work with a progressive organization.

Ambitious

The MCFA looks set to continue growing. Today, when Marie Curie fellowships are awarded, individuals are automatically invited to join the association.

Members are ambitious for the organization over the next five years, wanting it to be a stronger voice in the link between academia and industry, and, says Vetere Arellano, a "strong collaborator" with the commission, taking part in FP6 as a "consultant in issues of training and mobility for researchers".

Ledoux is keen for the association to carry political weight with both the European Commission and research funding organizations. The MCFA, she says, provides a voice not only for Marie Curie fellows but also for young scientists in general.

This is not an idea with which everyone is comfortable. Some observers say the association cannot speak for all young scientists, but only for the elite who win Marie Curie fellowships. They also question whether becoming a voice for young scientists is appropriate for an organization funded by the European Commission for a different reason.

But the impetus that is building up behind the MCFA could be difficult to stop. And the question remains: if the MCFA does not become the voice for young European scientists, who will?

Natasha Loder

Germany's research and education spend still set to rise

Munich German Chancellor Gerhard Schröder has confirmed that the budgets for research and education in Germany will increase in coming years. But he told last week's fiftieth anniversary meeting of the Fraunhofer Society, Germany's main organization for applied research, that a promised doubling of research spending "may take longer than intended".

Next year, the research budget will be exempt from a 7.4 per cent cut in government spending. But Schröder announced that annual funding of the Fraunhofer Society — currently DM468 million (US\$251 million) — will rise by three per cent in 2000 rather than two per cent, as in the past few years.

Area under transgenic crops shoots up 44 per cent

Tokyo The area of land planted with genetically modified (GM) crops is expected to increase dramatically this year, particularly in China, Argentina, Canada and South Africa, according to Monsanto, the US agri-biotechnology company.

Paul Teng, the company's director for

biotechnology in the Asia-Pacific region, said last week that almost 40 million hectares will be planted with GM crops this year, an increase of 44 per cent. Teng also said that GM crops were planted commercially in Portugal and Ukraine for the first time this year. Teng was speaking at the Asian rice conference in the Philippines.

South Africa names research chiefs

Cape Town Khotso Mokhele, 44, formerly head of South Africa's Foundation for Research Development, has been appointed head of the National Research Foundation, the new funding agency for research in natural, engineering and human sciences in higher education.

Rob Adam, 43, deputy director-general for science and technology in the Department of Arts, Culture, Science and Technology since 1996, has been made director-general of the department. Adam has good working relationships with his minister, Ben Ngubane, and researchers.

Appeal court deals blow to US environmental agency

Washington The US Environmental Protection Agency (EPA) suffered another legal set-

back last week when a US Court of Appeals let stand an earlier decision challenging the agency's ability to set clean-air regulations. Last May, a three-judge panel of the same court found that new EPA rules governing ozone and airborne soot were an "unconstitutional delegation of legislative authority".

In its 29 October split decision, the full court rejected EPA's request to reconsider the ruling. But agency administrator Carol Browner said in a statement that EPA will take the case to the US Supreme Court, where it will be in a "very strong legal position". Two of the 11 judges in the appeals court did not vote last week, and five sided with EPA in a dissenting verdict. This is one short of the number needed to grant the decision a further hearing.

US ecologists elect woman president

San Diego The Ecological Society of America has elected Diana Wall, a soil ecologist at Colorado State University in Fort Collins, as its president. Wall directs the Natural Resource Ecology Laboratory and is an associate dean of the College of Natural Resources at Colorado State.

As president of the 7,600-member society, Wall says that she is keen to make ecological knowledge more accessible to citizens and

decision-makers at all levels of government. "Americans depend upon ecological knowledge to manage our natural resources, protect human health and prevent small problems from becoming large, expensive problems," said Wall, who studies how diversity of invertebrate species makes soils productive.

Spain links up national quake detection system

Barcelona Spain is to apply advanced technology to its earthquake detection and warning system through a network that will link 40 monitoring stations nationwide within a year. The data registering system will become digitized, adding the north-south and east-west components to the current vertical one. Any earthquake registered in this way will connect the sensor to the reception centre by satellite, allowing 'real-time' detection. The scheme has a budget of Ptas 500 million (US\$3.3 million).

Emilio Carreño, head of the Spanish Seismologic Network, admits that the country has a relatively low rate of seismic activity compared to the rest of the world. However, he points out that the 1755 Lisbon earthquake, which registered 8.39 on the Richter scale, caused widespread damage to the Iberian peninsula.

Flood wreaks havoc at medical school

London A burst water main has caused extensive flooding and damage to buildings at Guy's Hospital in London, including a brand new multimillion-pound biomedical research and teaching centre, Hunt's House. Part of a major 36-inch trunk main is thought to have blown out in what water suppliers described as "a pretty dramatic" manner.

"It's one of the biggest bursts we've had to deal with in London this year," said a spokesperson. A complete inventory of damage is not yet available, but a spokesperson for Guy's Medical School confirms that the electron microscopy unit, containing two microscopes, has been lost. Some reports estimate damage at £2 million (US\$3.3 million).

Cloning company lays off staff in marketing setback

London PPL Therapeutics, the company holding a patent on the cloning technology used to create Dolly the sheep, is laying off staff and restructuring following a setback to its operations. PPL had hoped to secure a marketing partner to take its lead drug AAT — to control cystic fibrosis — into phase three trials.

PPL recently came close to signing a deal

with a major pharmaceutical company, but this was put on hold when the partner decided not to proceed immediately. PPL is now looking to reduce its rate of cash burn until a partner comes on board. Share prices on the London stock market for PPL are currently at 68 pence (US\$1.11), down from a high of 450 pence shortly after Dolly was cloned. PPL recently received a £2 million grant for its work on xenotransplantation (see *Nature* 401, 633; 1999).

Engineer Netravali to lead Bell Labs

Washington Arun Netravali, an electrical engineer and multimedia specialist, has been named as president of Bell Labs, the network of laboratories which are now the research arm of Lucent Technologies. Netravali, 53 (below), was already head of research at Bell Labs and has worked there for 27 years.



With 25,000 employees at Murray Hill, New Jersey, and other sites around the world, Bell Labs is among the world's largest industrial research centres. Netravali has a doctorate in electrical engineering from Rice University, Houston, Texas.

Europe must grant crucial funds for biological research

Sir — The impact of genetics on our lives is now obvious to any reader of a daily newspaper. The prospect that Europe may refuse to fund the infrastructure that underpins all biological research is simply extraordinary (see pages 3–4 in this issue). Such research is essential for the exploitation of genetic and genomic research for the benefit of humankind. Aside from far-reaching benefits in medicine and health care, this science will bring advances to fields as diverse as agriculture, manufacturing and the environment.

Human health and welfare have the potential of great benefit from genetics research, largely because developments in large-scale DNA sequencing now make it possible to determine the complete genome sequence of any living organism. The complete genome sequences of more than 20 prokaryotes and several 'model' organisms are now public, or will be so within a year — two yeasts, a nematode worm, the fruitfly *Drosophila* and the plant *Arabidopsis*. The sequencing of the mouse genome has begun and, most dramatically, a draft of the human sequence will be finished within six months — a singular event in human history.

This information is revolutionizing biological research. Its analysis, both computational and experimental, will give an extraordinarily deep understanding of living organisms. The use of this knowledge will have an impact on all our lives.

Much of this work has been done in European laboratories — traditionally strong in biology and genetics research. European researchers are poised to make major contributions in the future, many through collaborations between researchers in different European countries. A decade or so ago, biological research was small in scale and required relatively little in the way of infrastructure. Today, particularly in genomics, it is quite different, requiring large-scale, often collaborative, research underpinned by robust infrastructure.

Infrastructure is defined by the European Commission (EC) as "facilities and resources that provide essential services to the research community in the life sciences"¹. The paradox is that, although the essential nature of such infrastructure is recognized by the EC, particularly for bioinformatics and stock centres for model organisms, it is ineligible for funding under the commission's fifth Framework programme of research: "[it] will not

provide for support for tasks that involve the construction and routine operation of research infrastructures, nor for the collection of data...". This decision does not reflect the needs of the scientific community.

Although recognizing the importance of this infrastructure, the EC appears to be of the view that it should be funded by some other agency, or by individual countries. There is no other pan-European agency that currently has the ability to fund these "essential services". It is clearly inappropriate for infrastructure needed by all European scientists, in academic institutions and in industry, to be funded at national level.

A failure of funding now is particularly ironic given that we are on the verge of elucidating the entire human genome sequence. This prospect is a stark contrast with the commitment and vision shown by biology funding agencies in the United States and Japan. Because the decision not to make EC funding available for sustained infrastructure appears to be a political one, a political resolution must be found, and found soon. If not, then European research in the biological sciences, and all that flows from it, will suffer.

Michael Ashburner*

*Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK
e-mail: infrastructure@gen.cam.ac.uk*

*Readers can add their support to this letter by sending their name, title and affiliation by e-mail to M. A., the coordinating author. Full affiliations and addresses of the signatories below are in Supplementary Information on *Nature's* World-Wide Web site (<http://www.nature.com>) or can be obtained from M. A.

Paolo Amati, University of Rome; **Mike Bevan**, John Innes Centre, Norwich; **Anton Berns**, The Netherlands Cancer Institute; **Tom Blundell**, University of Cambridge; **Piet Borst**, The Netherlands Cancer Institute; **Graham Cameron**, EMBL — EBI; **Luca Cavalli-Sforza**, Stanford University; **Julio E. Celis**, Danish Centre for Human Genome Research; **Pierre Chambon**, IGBMC Strasbourg; **Antonio Coutinho**, Gulbenkian Foundation for Science, Portugal; **Antoine Danchin**, Institut Pasteur, Paris; **Bertil Daneholt**, Karolinska Institutet, Stockholm; **Jacques Demaille**, Human Genetics Institute, Montpellier; **Maria De Sousa**, Oporto University, Portugal; **T. Michael Dexter**, Wellcome Trust, London; **Denis Duboule**, University of Geneva; **Marvin Edelman**, Weizmann Institute of Science, Israel; **Arturo Falaschi**, International Center for Genetic Engineering and Biotechnology, Trieste; **Frank Gannon**, EMBO; **Antonio García-Bellido**, Universidad Autónoma de Madrid; **F. García-Olmedo**, Universidad Politécnica de Madrid; **Walter Gehring**, University of Basel; **Peter N. Goodfellow**, SmithKline Beecham; **Frank Grosveld**, Erasmus University, Rotterdam; **Peter Gruss**, Max-Planck-Institut für Biophysikalische Chemie, Göttingen; **Bernhard Hirt**, Swiss Institute for Experimental Cancer Research, Epalinges; **Louise Johnson**, University of Oxford; **Alwyn Jones**, Uppsala University; **Bertrand R. Jordan**, CIML Luminy, Marseille; **Fotis Kafatos**, EMBL; **Kari I. Kivirikko**, Academy of Finland; **Jonathan Knowles**, Hoffman LaRoche; **Nicole le Douarin**, Collège de France, Paris;

Mark Lathrop, Centre National de Genotypage, Evry; **Chris Leaver**, University of Oxford; **Marja Makarow**, Universities of Kuopio and Helsinki; **Bernard Malissen**, Marseille Immunology Centre; **Matthias Mann**, University of Southern Denmark; **David McConnell**, Trinity College Dublin; **Juan Modolell**, Centro de Biología Molecular Severo Ochoa, Madrid; **Christiane Nusslein-Volhard**, Max-Planck-Institut für Entwicklungsbiologie, Tübingen; **Mary Osborn**, Max-Planck Institut für Biophysikalische Chemie, Göttingen; **Lennart Philipson**, Karolinska Institute; **Ronald Plasterk**, Hubrecht Laboratory, Utrecht; **Hans Prydz**, University of Oslo; **Francis Quetier**, Genoscope, Evry; **Klaus Rajewsky**, Institute of Genetics, Cologne; **Peter W. J. Rigby**, Institute of Cancer Research, London; **Cecilia Saccone**, CNR Research Area, Bari, Italy; **Gottfried Schatz**, Biozentrum, University of Basel; **Piotr Slonimski**, CNRS, Gif-sur-Yvette; **Ed Southern**, University of Oxford; **John Sulston**, Sanger Centre, UK; **Joel L. Sussman**, Weizmann Institute of Science, Israel; **George Thireos**, IMBB, Crete; **Glauro Tocchini-Valentini**, IBC-CNR, Monterotondo Scalo, Italy; **Peter van der Vliet**, Centre for Biomedical Genetics, The Netherlands; **Marc Van Montagu**, University of Gent; **Erhard Wintersberger**, University of Vienna; **Moshe Yaniv**, Institut Pasteur, Paris; **Rolf Zinkernagel**, University of Zurich; **Harald zur Hausen**, Deutsche Krebsforschungszentrum, Heidelberg.

1. Quality of Life Programme Management, European Commission Support for Research Infrastructures (EC, Brussels, 11 October 1999).

University becomes political football

Sir — Alejandro Cuevas-Sosa distorts the picture of what is happening at the National Autonomous University of Mexico (UNAM) (*Nature* **401**, 524; 1999). UNAM has been closed for almost six months owing to a strike by students against new regulations for fees that were virtually non-existent. The academic staff have not been on strike, contrary to what Cuevas-Sosa says. Moreover, research in institutes is continuing, given that we have limited access to our laboratories in the main campus in Mexico City. Undergraduate teaching has in many cases stopped and is happening only partially in other places.

The academic arguments that started this lengthy conflict are no longer valid. The dispute has become a political argument over the most important cultural and academic project this country has ever supported. The opinions of those of us who carry out full-time academic activities in this university have been ignored or dismissed.

I raise a plea for support by scientists around the world to fight against using universities as political arenas. Governments should be alerted to the importance that research and teaching has had worldwide and to the beneficial influence these activities have on society.

Georges Dreyfus

*Instituto de Fisiología Celular,
Universidad Nacional Autónoma de México,
Ap Postal 70-243, 04510 México DF, México*

Making sense, making money

An enlightened approach to using energy and materials.

Natural Capitalism: The Next Industrial Revolution

by Paul Hawken, Amory B. Lovins and L. Hunter Lovins

Earthscan: 1999. 396 pp. £18.99

Norman Myers

“Oil will become uncompetitive even at low prices before it becomes unavailable even at high prices.” So says Amory Lovins, one of the authors of a visionary book that is truly appropriate for the new millennium with its focus on ultra-efficiency of energy and materials. Lovins and his two co-authors postulate that we need to adopt a radical new approach to the way we live with the Earth. The Industrial Revolution enabled us to mobilize natural resources without restraint. Now that natural resources — not just traditional materials and energy sources, but water, topsoil, forests and the climate itself — are being degraded and depleted, we must move towards a capitalism that safeguards the natural resource base that ultimately underpins all economic activities. As the book says: “Nature bats last and owns the stadium.”

Conventional capitalism is remarkably inefficient. In the United States, materials used by industry’s metabolism every day amount to more than 20 times the total weight of the American population. Yet only one per cent of these materials ends up in products that are still in use six months after being sold, the rest being junked. It would be better to have industrial parks where each manufacturer feeds upon the wastes of others until emissions are finally reduced to zero. This ‘industrial ecology’ eliminates not only waste, but the concept of waste itself — as is practised by nature with its closed-loop ecosystems.

There are success stories along these lines in Sweden, Denmark, Colombia, Namibia, Kenya and Fiji. The Ebara Corporation in Japan and DuPont in the United States are formally committed to zero waste, and the Asahi Beer Company in Tokyo already recycles 98.5 per cent of its raw materials. Bristol-Myers Squibb claims that the economic benefits of pollution prevention exceed the costs fourfold.

Much the same applies to energy efficiency. The US Energy Star Program enables televisions and videos to achieve a 75 to 95 per cent reduction in the energy used in stand-by mode, which currently costs Americans more than \$1 billion a year.

To achieve sustainable economies, we need to reduce our materials and energy intensity (the amount used per unit prod-

uct) by 50 per cent worldwide. Given that developing countries will be reluctant, let alone unable, to do this for some time, developed nations should aim to cut theirs by 90 per cent. Through numerous examples, the book shows how this can be done while enhancing our quality of life. It also shows why natural capitalism will often generate super profits.

A 90 per cent reduction should not be as difficult to achieve as it might appear. When we learned to substitute coal, machines and technologies for human muscle, we expanded worker productivity 200-fold within half a century. In similar though less significant style, German manufacturers now find they can dispense with 98 per cent of ‘secondary’ packaging around toothpaste tubes, ice-cream cartons and the like.

The strategy of a 90 per cent reduction in materials is known as ‘factor 10’. It is entering the vocabulary of government officials, economic planners, scientists and business leaders around the world. The governments of Austria, the Netherlands and Norway have already committed themselves to pursuing 75 per cent, or factor 4, efficiencies, which means that, by using existing technologies, we could enjoy twice as much well-being while using half as many materials and causing half as much waste.

The same strategy has been endorsed by the European Union as the new paradigm for sustainable economies. Better still, Austria, Sweden and the OECD (the Organization for Economic Co-operation and Development) have urged the adoption of factor 10, as has the World Business Council for Sustainable



Making no sense, making money

Man’s abuse of the planet’s biological resources is vividly illustrated by this image, one of 143 pictures featured in the BG Wildlife Photographer of the Year Exhibition at the Natural History Museum, London.

Winner of the category entitled ‘The World in Our Hands’, this powerful image depicts a slaughtered family of lowland gorillas in Cameroon. The photographer, Karl Ammann, explains that the gorillas were cornered, chased up trees by trained dogs and then shot down. By persuading the hunters to arrange their spoils at the foot of a tree he was able to record the senseless killing. The hunters would probably

receive just £30 for the adults. And the babies? Local children would be given them to play with and eat.

The exhibition runs in London until 27 February 2000, with travelling exhibitions touring Britain until February 2001 and an international tour visiting countries including Australia, Germany and the United States. It features a selection of the best photographs in 15 categories, ranging from ‘The Underwater World’ to ‘Urban and Garden Wildlife’. The overall winner, ‘Leopard with Rising Moon’ by Jamie Thom, shows a leopard resting at dusk, gazing lazily into the camera. *Alison Mitchell*

Development. Leading corporations such as Dow Europe and Mitsubishi Electric see it as a powerful approach for gaining competitive advantage.

Illustrative of the radical new approach is the so-called hypercar conceived by Amory Lovins. Because it would be largely made of advanced polymer composites, especially carbon fibre, it would use one-third less aluminium, three-fifths less rubber, four-fifths less platinum and nine-tenths less steel, and so would weigh only one-third as much, as the cars of today. Thanks to these and other design efficiencies, notably a hybrid-electric drive, it would achieve 140 to 250 miles per imperial gallon, it would be 95 per cent less polluting, and it would be almost entirely recyclable. Hypercars could eventually save as much oil as OPEC (the Organization of Oil Producing Countries) now sells. They would be not so much cars with microchips as computers on wheels.

Other 'techno-breakthroughs' already in view include: diodes that emit light for 20 years without bulbs; ultrasound washing machines that use no water, heat or soap; deprintable and reprintable paper; plastics that are both reusable and compostable; roofs and roads that double up as solar-energy collectors; extra-light materials that are stronger than steel; and quantum semiconductors that store vast amounts of information on chips no bigger than a dot. Fortunately, technological advances are reaching the marketplace faster and faster. In the United States, electricity was taken up by one-quarter of the population in 46 years; the telephone needed 35 years; the microwave oven, 30 years; television, 26 years; radio, 22 years; the personal computer, 16 years; the mobile phone, 13 years; and the World-Wide Web, just 7 years.

We could also derive far greater benefit from biological resources, which so far have been over-exploited and under-utilized. Through what is known in the trade as 'biomimicry', we could draw design inspiration from spiders which, using flies as feedstocks, make silk as strong as Kevlar yet much tougher, and without needing high-temperature extruders. The abalone manufactures an inner shell twice as tough as the best ceramics and diatoms make glass, both processes using sea water and no furnaces.

Over-visionary as some of the book's findings may appear to some eyes, the advance towards natural capitalism is surely as inevitable as it is possible. We live at a time when, by force of environmental circumstance (which is becoming ever more forceful), there is a strong convergence between the idealistic and the realistic. For empirical evidence, check the book's hundreds of "insurmountable opportunities".

We have hitherto sought to exploit the resources of the planet in support of the human cause. Now we need to exploit a

human resource — brain power — in support of the planetary cause, and thereby give ourselves an expanded prospect of securing the human cause as well.

Of course, there is a consumerist risk in all this. For example, if the hypercar is so cheap to drive, won't we end up with three or more cars in every garage? And the same with the other 'techno-innovations', leaving us with an overloaded planet once again? To this, Lovins offers another aphorism: "More is better, enough is best." This would be a social revolution indeed — and it would be in accord with the spirit of this inspiring book. ■

Norman Myers is at Green College, Woodstock Road, University of Oxford, Oxford OX2 6HG, UK.

Condensed physics matters

More things in Heaven and Earth: A Celebration of Physics at the Millennium

edited by Benjamin Bederson
American Institute of Physics: 1999. 600 pp.
\$78, £61

Philip W. Anderson

The cliché about designing things by committee is true. Some appropriate body — perhaps the Council of the American Physical Society (APS) — decided that an issue of *Reviews of Modern Physics* should celebrate the millennium and the society's centennial. Somebody else decided that it would be nice to put the result between hard covers. Marty Blume, editor-in-chief for the APS, and Ben Bederson, editor of *Reviews*, appointed a committee of six very fine physicists, each representing one of the major fields of physics, who did the best they could. In most cases they produced major articles under their own authorship as well as selecting a fine group of colleagues.

So why did I come away dissatisfied? For one thing, the same organizational procedure has been used every ten years or so by the National Academy of Sciences, and issued as a report on the state of physics in the United States (the Brinkman report, the Pake report ...). None of these are best sellers, although they are remarkably useful and in parts well written. What is wrong with them? Aside from fragmentation, there is the inevitable competitiveness which results in each section being written as though we have just reached a pinnacle of understanding. There can be no hint that behind the present synthesis there may lie decades of crisis, misunderstanding and even bitter controversy, which are unlikely to have been resolved just at this moment. There is also little sense of the essential unity of physics, in terms not of sub-

ject matter but of much more important but less tangible things such as mathematical structure and epistemology.

The present volume is an improvement in that there are several articles of a straightforward historical or autobiographical character, and to me these were by far the best part of the book. Many, but for some reason not all, of these are assembled in an initial section called "Historic Perspectives".

This section contains a number of gems: Hans Bethe on the early history of the quantum theory, Bram Pais on the history of theoretical particle physics, Val Fitch fascinatingly recounting the emulsion period of particle experiment, and Bob Pound's personal story of the early days of NMR, were the most rewarding. In the specialist sections, there was a notable article on the laser by Willis Lamb, Charles Townes *et al.* in much the same anecdotal vein.

The book then proceeds with separate sections on the various fields of physics, much in the tradition of the Pake report. It even follows the inescapable order, affording pride of place to particle physics and sliding down the energy scale (and up, may I say, the scale of relevance) until, concealed in the back of the book, we come to biophysics.

As one has come to expect, the particle physicists make their case with great clarity and intelligence. Particularly lucid and enlightening was Frank Wilczek's survey of field theory.

The astrophysicists perform according to form, opening with a dense technical article by Tyson and Turner which to me sounded particularly vulnerable to the experimentally driven revisions characteristic of cosmology. (I have pinned on my office door the *New Yorker* cartoon whose caption is: "Scientists today confirmed that everything we know about the universe is wronged-wrong-wrong.") But the articles on black holes, the microwave background and cosmic rays contain so much vital material that the net result can hardly fail to be exciting. The section ends, however, with not one but two articles devoted to seemingly hopeless searches: for dark matter and gravitational radiation. In a book where so much fascinating physics has been skimmed, perhaps these could have been omitted? Or better yet, replaced with a serious discussion of the politics, economics and sociology of such large experimental devices.

I can't cover all the sections in such detail. In a period when nuclear physics is reinventing itself as a mixture of applied quantum chromodynamics and astrophysics, perhaps I can skip over this; and atomic physics seemed to be focused on current, if interesting, news from the research front.

When we come to condensed matter and statistical physics, which I know a bit about, it is not at all clear where the dividing line between them lies: are granules less con-



An eye for physical detail: a lab technician inspects the revolutionary transistor back in 1952.

densed than superfluids? It might have been wiser to use the style current in the field, 'soft' versus 'hard' condensed matter.

In the end, the responsibility for the inadequate coverage of condensed matter lies with the editor: he has not given enough space to the enormous scope of this field, which employs half of all physicists. His prejudices are showing. Particle physics and astrophysics are given 100 pages apiece, whereas condensed matter has a scant 50. Each of the other major subjects is at least permitted a lead article of some general coverage and depth. As it is, Walter Kohn's article carries an impossible double burden — to summarize the history and set the intellectual background for the specialized articles that follow.

This latter task it does not do: there is no mention, for instance, of such broad principles as broken symmetry or the Mott phenomenon. The six brief specialized articles are left to fend for themselves. There is an excellent piece on the recently laurelled quantum Hall effect by Stormer, Tsui and Gossard, and a nice exposition of the miraculous new scanning microscopy by Binnig and Rohrer. The rest are almost uniformly perfunctory: Schrieffer and Tinkham fail to mention cuprates in their chapter on superconductivity, much less heavy electrons and organics; Leggett, on superfluidity, ignores ^3He , to which he contributed so much. That makes three recent Nobel prizes not mentioned in the index. Materials physics merits only six very brief pages by Chaudhuri and Dressel-

haus. I can only suppose that some of the authors gave up in disgust at the restricted compass they were allowed, and/or quailed at the possibility of portraying contention and uncertainty, which is, after all, the lifeblood of physics as it actually happens.

The excellent section on statistical physics restores some balance — at least, the great principles of scaling, universality and renormalization get a full treatment by Stanley, and a sequence of shorter articles introduce some of the main topics of modern soft physics, with an excellent article on pattern formation by Jim Langer and Jerry Gollub.

The book closes with a nod towards biophysics and a section on technology; the piece by Sid Drell on physics and security is informative and well written, standing out among a sequence of good but brief contributions.

This book attempts the impossible, starting off towards many destinations, but reaching few or none. One questions whether it is meaningful to mix serious scientific reviews with anecdotal material, seasoned with time-serving 'surveys' on the level of a textbook or government report. And there seems a massive distortion of emphasis in favour of journalistically popular fields. It has many articles that will be useful in a library. As for a more general audience, I can't recommend it. ■

Philip W. Anderson is at the Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA.

Simple twists of fate

Chance, Development and Aging

by Caleb E. Finch and Tom Kirkwood
Oxford University Press: 1999. 258 pp.
£24.95, \$35

Jonathan Hodgkin

Chance is central to the lives of organisms. Birth, reproduction and death all involve encounters with random external events, and so it is scarcely surprising that exact predictions of an individual's lifespan and fertility should be impossible. Random internal events are equally unavoidable, because all biological reactions take place in a sea of thermal fluctuations. As a result, no two cells can ever be exactly alike. In a memorable 1976 experiment, John Spudich and Daniel Koshland examined sets of isogenic bacteria for their chemotactic behaviour, tracking them for more than an hour. Different cells exhibited distinct distributions of 'runs' and 'tumbles', and retained these characteristics over the period of a lifespan. So even genetically identical bacteria have individual character, presumably assigned to them by chance fluctuations in the numbers and states of a few key regulatory molecules.

Caleb Finch and Tom Kirkwood have put

together a book which suggests that chance is under-appreciated as a factor in development and life history. However, the importance of chance is abundantly clear to any developmental biologist. The eggs of many animals start out as symmetrical objects that depend on symmetry-breaking events to generate the major embryonic axes, often relying on stochastic events to achieve this. Chance continues to operate as fields of equipotential cells undergo differentiation and commitment to different fates.

Chance, Development and Aging has a fairly misleading title. It begins with a sensible discussion of variability in ageing and lifespan, but then gets lost in reviews of mammalian reproductive biology and developmental neurobiology, which take up a large part of the text. These are interesting enough but are often of limited relevance. In the context of the book, the purpose of these surveys should be to assess how much chance, as opposed to genotype and external environment, can influence life-history traits such as fecundity and lifespan. This can really only be done in studies involving monozygotic twins, or highly inbred populations of laboratory animals such as mice, flies and worms. Yet we are given much information from studies of general populations, some of which is totally irrelevant — for example, human gender differences in acoustic physiology are illustrated and discussed.

The human twin studies discussed are undoubtedly informative. It is valuable to document the differences in brain anatomy between identical twins, even (in one pair) to the extent of a demonstrable difference in the side of the brain used to control swallowing. Twin pairs exhibit considerable discordance in Alzheimer's disease, spinal-disk degeneration and overall lifespan. But again, this is not surprising. To an immunologist, it will be axiomatic that no two individuals are identical. The adaptive immune system uses chance as an essential element in generating a diverse immune repertoire. Probably the nervous system goes through equivalent processes as it becomes wired up, so the final connectivities will be endlessly variable between individuals, however similar their genotypes.

Finch and Kirkwood are on firmer ground in discussing simple, inbred invertebrates. It is impressive that populations of the nematode *Caenorhabditis elegans*, effectively isogenic and homozygous at all loci, nevertheless exhibit considerable individual variations in lifespan. But the authors are overly impressed by the defined lineage of this creature. Most adult *C. elegans* do contain precisely 959 somatic nuclei, but the underlying lineages contain many binary choice points with respect to developmental fate, so the chance of two worms having truly identical lineages is very small. For all we

know, the precise lineage details may significantly affect each individual's prospects.

In fact, the miracle is that isogenic animals are as similar as they are. Developmental canalization is a very powerful force, which becomes apparent when normal development is perturbed by experiment or mutation. A clone of animals carrying the same single deleterious mutation will sometimes exhibit startling variability in phenotype, while the parental line remains conspicuously uniform. This uniformity is the result of canalization: developmental processes use many error-correcting devices, operating at every level from enzymes to whole organ systems, in order to create highly invariant final structures. The book touches on error-correcting mechanisms, but does not treat them extensively, nor adequately consider what factors may limit their operation. In some situations, as in the immune system, chance may even be retained as a selectively advantageous factor.

Conversely, there are all kinds of possible epigenetic mechanisms that can turn an initial fluctuation into a long-lasting and even permanent change in the life of a cell or an organism. Finch and Kirkwood refer briefly to the possible role of mitochondrial variation as an epigenetic mechanism, but do not explore other possibilities. Their timing is unfortunate, as they discuss the famous cloned sheep Dolly without referring to the recent demonstration that she has a different set of mitochondria from the udder cell that contributed her nuclear genome.

The book ends with various recommendations. One of these is that the familiar partitioning of phenotypic variance into one part attributed to genotype and one part to environment should be extended to include a third part attributed to chance. However, it will be a tall order to separate the contributions of environment and chance. Most biologists will be content simply to remember that 'environment' will always include some uncontrolled variables.

Much of the book's remaining agenda concerns general current research on development and ageing. Perhaps their most salient point is to emphasize the importance of very early developmental events, some of them probably due to chance, in determining the life-history traits of long-lived animals such as humans. To this end, the authors are keen to see the development of technologies for assessing variation in organs such as ovary and prostate at birth, since these seem to have significant long-term consequences for fertility and lifespan. Overall, *Chance, Development and Aging* contains much interesting material, but it could have been distinctly better. ■

Jonathan Hodgkin is at the MRC Laboratory of Molecular Biology, MRC Centre, University Medical School, Hills Road, Cambridge CB2 2QH, UK.

Science in culture

Wozzeck, an opera by Alban Berg based on the prophetic play by Georg Büchner

John Carmody

Science is not the only aspect of human creativity that seeks to be experimental and prescient. This ought to be an enduring aspiration of art, too. And opera — notwithstanding Dr Johnson's denigration of it as "an exotic and irrational entertainment" — has that capacity in full measure. Mostly, though, opera has neglected science and trivialized medicine. But there is one chilling exception, that twentieth-century masterpiece, Alban Berg's *Wozzeck*. It was based on Georg Büchner's prophetic play of 1835–37, a work that was the more telling because it was loosely based on real events.

Büchner, the son of a Darmstadt doctor, graduated in medicine in 1834 but was as dedicated to social and literary reform as to medicine: had he not died of typhus in 1837, while a young lecturer in comparative anatomy in Zürich, he may well have outstripped Goethe in literary achievement. His *Wozzeck* (or *Woyzeck*, in the manuscript of the play) is the model of the anti-hero we know so well in modern literature, the archetype of the hapless, poverty-stricken common man; in particular, he is exploited by medical science. There is, therefore, a timeliness in the new production, under the illuminating direction of Barrie Kosky, which has been at the Sydney Opera House — we have, again, a vantage point for reflection on the role of science and medicine in this century, our retrospect to complement Büchner's prospect.

"We poor people," *Wozzeck* laments to his regimental captain. "If I had a hat, a watch and an eyeglass ... I would be virtuous, too. Folk like us are always unfortunate in this world." Unfortunate, indeed for *Wozzeck*, forced by the need to support his bastard child to participate in the town doctor's bizarre dietetic experiments, to be the victim of a *parvenu* scientist who likens *Wozzeck* to a lizard. The doctor is irritated that this benighted Everyman — a mere experimental subject — is not strictly adhering to his unvaried diet of beans and, worse, urinates in the street rather than into the doctor's collection flasks.

The doctor is hostile to the notion of an individual nature. "Mere superstition. Have I not proved that the diaphragm is subject to the Will? Individuality is sublimated into freedom." Have there been repeated echoes (from scientists and doctors) of that philosophy in our century? The doctor, obsessed with his own importance — "Oh, my hypothesis! Oh, my fame!" — glibly redefines *Wozzeck*'s yearning to express his humanity as a psychiatric disorder, "an excellent *aberratio mentalis partialis*, second species", and later he reveals himself to be equally callous with his patients.

Here we have an allegory of what science would do to our world, not in a detached way,



through weaponry and technology, but face to face with people — the betrayal of medical and scientific trust in the horrible experiments of the Nazi era; the complicity of doctors in torture and murder (and not only in Germany, as Neil Bolton's recent biography, *The Good Listener — Helen Bamber: A Life Against Cruelty* [Weidenfeld & Nicolson], reminded us); the perfidy of medicine in the abuse of psychiatry for so many years in the Soviet Union. It is no exaggeration to say that they are all presaged in Büchner's play; he was describing inhumanity to come, abuse of hapless Everyman and Everywoman on an unprecedented scale.

The extra twist in *Wozzeck* — and Berg's colourful, controlled music makes it more telling — is that the doctor's parasitic indifference to the human values that ought to underpin science and medicine applies to the higher levels of his society no less than to the common soldiers and their women. The doctor almost gloats at the prospect when he warns his friend, the captain, "You might well have an *apoplexia cerebri* someday ... I can assure you that it will be a most interesting case. If God wills it, your tongue will be partially paralysed and we'll be able to do immortal experiments."

In such a society — in *our* society, which Büchner foresaw — every stratum exploits whatever lies beneath it. If art — opera most potently — seeks to "hold the mirror up to nature", *Wozzeck* should compel all of us scientists to look deeply into that reflection of our true souls. ■

John Carmody is in the Faculty of Medicine, University of New South Wales, Sydney 2052, Australia.

Are you serious, Dr Mitchell?

Few would have laid money on cells generating energy with proton pumps.

Leslie E. Orgel

The development of molecular biology since its spectacular beginnings in the 1950s has provided the informed participant or onlooker with a series of breakthroughs and surprises. However, few of the insights would have tempted workers in the same field to respond with the question "Are you serious?" In this respect the history of biology is different from that of physics, which has included a series of assaults on 'common sense'. It is alleged that distinguished Oxford professors urged their students to ignore relativity theory because it would soon go away. I doubt that many people felt that way about the structure of haemoglobin or DNA. Not since Darwin and Wallace has biology come up with an idea as counterintuitive as those of, say, Einstein, Heisenberg and Schrödinger.

The tradition of molecular biology owes much to the influence of Max Delbrück: formulate precise alternative theories of the phenomenon under discussion, and then design experiments to distinguish between them; but moving too far ahead of the facts is to be discouraged. This approach works wonderfully in biology where, as yet, few surprising general principles have emerged.

One of its consequences, or perhaps one of the consequences of the nature of biology, is that surprising new discoveries are rapidly assimilated. When one first read about reverse transcriptases, introns or ribozymes, one might have wondered about their origins or functions, but one didn't doubt that they were here to stay. The experimental evidence was either already convincing or very soon became convincing.

There may be a few exceptions to this a posteriori transparency of advances in molecular biology but, for personal reasons, one stands out in my memory. In 1955 I took up a position in the Department of Theoretical Chemistry at the University of Cambridge. For the next few years, I worked on problems in inorganic chemistry but, much influenced by George Beadle's elementary course on experimental biology at the California Institute of Technology, I was already spending a good deal of time talking with the Cambridge molecular biologists. One day a young scientist unknown to me made an appointment to talk about a theoretical matter that he thought would interest me. He was Peter Mitchell, and he wanted to talk about his ideas on how living cells derive energy: his novel chemiosmotic hypothesis (see *Nature* **191**, 144–148; 1961 and, for a

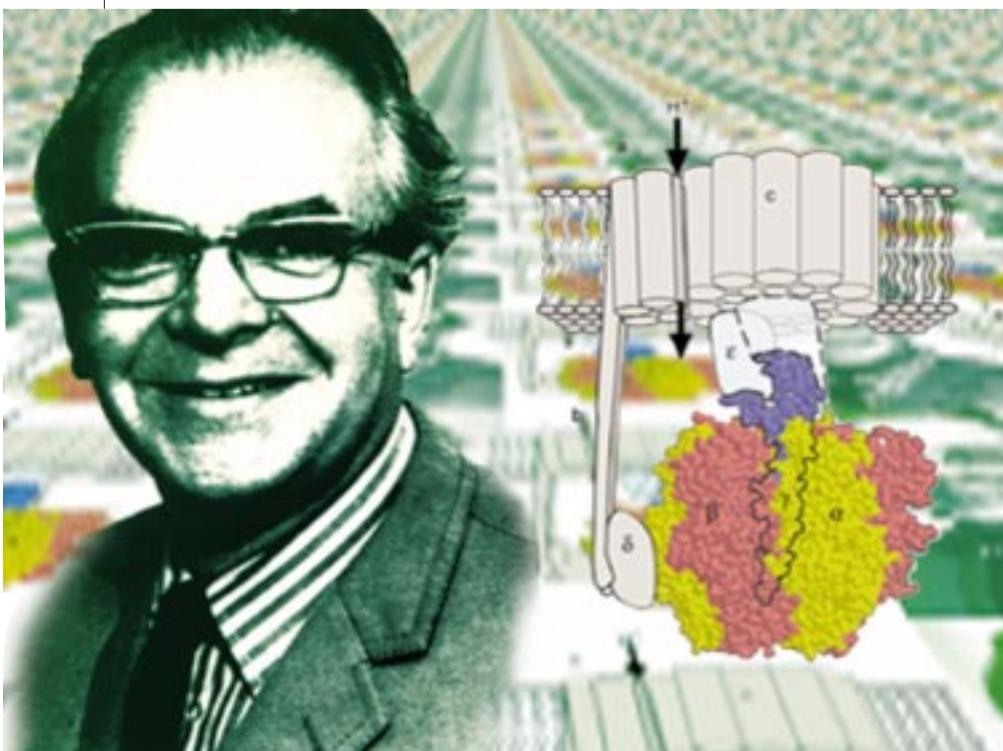
Not since Darwin has biology come up with an idea as counterintuitive as those of, say, Einstein.

review, Saraste, M. *Science* **283**, 1488–1493; 1999). According to Mitchell's ideas, metabolic energy was used to pump protons across a biological membrane thus establishing a concentration gradient. It was the return of protons down the gradient that led to the synthesis of ATP.

I don't remember the exact date or the details of our conversation, but I do recall its general trend. We discussed in some detail the energy that could, in principle, be obtained by transporting an ion across a membrane from a more concentrated to a less concentrated solution. I don't remember whether we took account of the electrostatic potential. I confirmed Peter's calculation that there might be enough energy available from the transport of protons across biological membranes to generate 'high-energy' phosphate bonds. I was too polite to express a view on the likelihood of Peter's mechanism being correct, but I remember thinking to myself that I would bet anything that ATP synthesis didn't work that way. Fortunately, no one took my bet.

At the time, most people familiar with the problem took it for granted that the formation of ATP would involve the reaction of ADP with a covalent phosphate-containing intermediate. I have no direct knowledge of the attitude of the biochemists, but I know that most molecular biologists shared my scepticism about Peter's ideas. I am not suggesting that his work should be compared to that of Darwin or the great physicists, but it did involve a paradigm shift and his ideas seemed bizarre to most of his contemporaries. They might well have asked "Are you serious, Dr Mitchell?" Of course, he was and he was right. Peter funded himself for much of his subsequent work, so we will never know how his research would have fared at the hands of granting agencies. I wouldn't count on much success in comparable circumstances at the beginning of the next millennium!

Leslie E. Orgel is at the Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.



He was serious: Mitchell and ATP synthase, which pumps protons across cell membranes.

Improving the neighbourhood

The death of a nearby star system comes as a relief — and a warning.

Arthur C. Clarke

At last, after feats of information processing that taxed our resources to the limit, we have solved the long-standing mystery of the Double Nova. Even now, we have interpreted only a small fraction of the radio and optical messages from the culture that perished so spectacularly, but the main facts — astonishing though they are — seem beyond dispute.

Our late neighbours evolved on a world much like our own planet, at such a distance from its sun that water was normally liquid. After a long period of barbarism, they began to develop technologies using readily available materials and sources of energy. Their first machines — like ours — depended on chemical reactions involving the elements hydrogen, carbon and oxygen.

Inevitably, they constructed vehicles for moving on land and sea, as well as through the atmosphere and out into space. After discovering electricity, they quickly developed telecommunications devices, including the radio transmitters that first alerted us to their existence. Although the moving images these provided revealed their appearance and behaviour, most of our understanding of their history and eventual fate has been derived from the complex symbols that they used to record information.

Shortly before the end, they encountered an energy crisis, partly triggered by their enormous physical size and violent activity. For a while, the widespread use of uranium fission and hydrogen fusion postponed the inevitable. Then, driven by necessity, they made desperate attempts to find superior alternatives. After several false starts, involving low-temperature nuclear reactions of scientific interest but no practical value, they succeeded in tapping the quantum fluctuations that occur at the very foundations of space-time. This gave them access to a virtually infinite source of energy.

What happened next is still a matter of conjecture. It may have been an industrial accident, or an attempt by one of their many competing organizations to gain advantage over another. In any event, by mishandling the ultimate forces of the Universe, they triggered a cataclysm which detonated their own planet — and, very shortly afterwards, its single large moon.

Although the annihilation of any intelligent beings should be deplored, it is impossible to feel much regret in this particular case. The history of these huge creatures contains



OLIVER BURSTON

Although the annihilation of any intelligent beings should be deplored, it is impossible to feel much regret in this case.

countless episodes of violence, against their own species and the numerous others that occupied their planet. Whether they would have made the necessary transition — as we did, ages ago — from carbon- to germanium-based consciousness, has been the subject of much debate. It is quite surprising what they were able to achieve, as massive individual entities exchanging information

at a pitifully low data rate — often by very short-range vibrations in their atmosphere!

They were apparently on the verge of developing the necessary technology that would have allowed them to abandon their clumsy, chemically fuelled bodies and thus achieve multiple connectivity: had they succeeded, they might well have been a serious danger to all the civilizations of our Local Cluster.

Let us ensure that such a situation never arises again. ■

Dedicated to Drs Pons and Fleischmann, Nobel laureates of the twenty-first century.

© Sir Arthur C. Clarke 1999.

Sir Arthur C. Clarke is chancellor of the International Space University and the University of Moratuwa, Sri Lanka. He is the author of 2001: A Space Odyssey and many other novels and stories, and was nominated for the Nobel Peace Prize for inventing the communications satellite. His latest book is Greetings, Carbon-Based Biped! He lives in Sri Lanka.

Catching the first fish

Philippe Janvier

Most major animal groups appear suddenly in the fossil record 550 million years ago, but vertebrates have been absent from this 'Big Bang' of life. Two fish-like animals from Early Cambrian rocks now fill this gap.

The search for the deep past of the vertebrates has been a pet subject of palaeontologists during the past decade. Traditionally, the fossil record of fishes and their descendants is good from the present to the Early Silurian period, about 430 million years (Myr) ago, then very poor in the Ordovician period (until about 480 Myr ago), and totally lacking or controversial before that, in the earliest Ordovician and Cambrian periods (from 480 to 550 Myr ago) and earlier. On page 42 of this issue, Shu *et al.*¹ report the discovery of two fish-like fossils from Chengjiang in Yunnan, China, that are probably the long-awaited Early Cambrian vertebrates.

British and Australian palaeontologists have found vertebrate remains from the beginning of the Ordovician². Ordovician vertebrates have been shown to be far more diverse than was previously thought, consisting of a mixture of primitive, jawless groups

and comparatively advanced groups with jaws³. However, all the Cambrian fossils thought to be vertebrates are from the Late Cambrian, and are controversial. They are chiefly the euconodonts (often called conodonts), a group long known from isolated denticles (small, tooth-like fossils) of debated derivation, but now known from complete, eel-like animals that are probably vertebrates⁴. Some palaeontologists dispute this, although they accept that euconodonts may be related to vertebrates^{5,6}.

Other putative Cambrian vertebrates are represented by small carapace fragments, whose microstructure recalls a group of jawless vertebrates with bony armour — the so-called ostracoderms — that lived from the Ordovician to the Devonian (Fig. 1). Again, some palaeontologists regard these fragments as belonging to true vertebrates⁷, whereas others think they are carapace fragments from arthropods related to horseshoe

crabs⁸. The exploration of the deep past of the vertebrates, before the Silurian period, is not for the faint-hearted, requiring the painstaking study of sketchy remains, with the constant danger of misinterpretation.

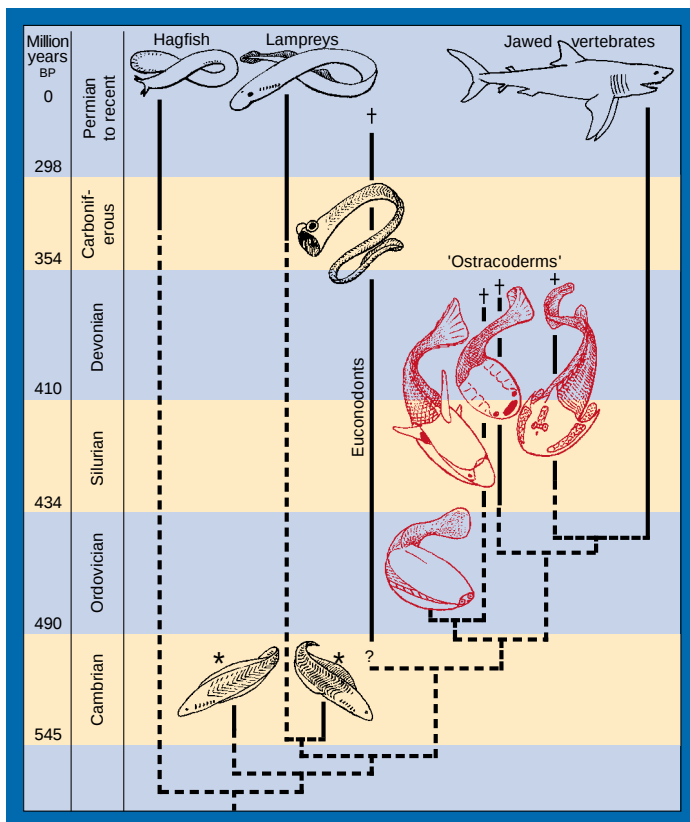
In fact, the bone-like fragments or denticles from the Cambrian might not lead to the 'root' of the vertebrates. The currently accepted phylogenetic trees showing the evolutionary relationships of fossil and living vertebrates suggest that mineralized tissues are a comparatively recent invention, and that the jawless ostracoderms are more closely related to the jawed vertebrates than to either lampreys or hagfish, the two living, jawless vertebrate groups, which have no bones or teeth (Fig. 1)⁹. This means that the earliest history of the vertebrates can only be documented by fossils formed under certain conditions, where the imprint of soft tissues has been preserved. For the Cambrian period, there are two famous sites where an exceptional preservation of the soft tissues has occurred. One is Burgess, in Canada, and the other is Chengjiang. Fossils resembling chordates (the large animal group that includes the vertebrates) or vertebrates have been recorded from both, such as *Pikaia* in Burgess, and *Yunnanozoon* and *Cathaymyrus* in Chengjiang^{10–12}. Although probably close relatives of the vertebrates, none of these fossils looks entirely familiar to a vertebrate specialist.

The two new fossils described by Shu *et al.*¹ from Chengjiang are the most convincing Early Cambrian vertebrates ever found. With their zigzag-shaped muscle blocks, relatively complex and presumably cartilaginous skull, gill arches, heart, and fin supports, these fish resemble the larvae of living lampreys (Box 1, overleaf). One detail, however, is at odds with what we know about vertebrates: the peculiar, forward tilt of the endoskeletal rays of the dorsal fin, which contrasts with the backward tilt of these fin rays in most fishes. Whether this is due to distortion or not awaits confirmation. Surprisingly, these animals seem to have long, ribbon-shaped, paired fins, which were thought to have appeared later in vertebrate history⁹, after the divergence of lampreys.

Shu *et al.* have analysed the phylogenetic position of these two Cambrian fossils in the framework of previous analyses of vertebrate phylogeny. Strangely, one of the two species seems to be more closely related to lampreys

Figure 1 The current phylogenetic tree of the major living and fossil vertebrate groups. This tree suggests that bone appeared relatively late, after the divergence of hagfishes and lampreys — the two living groups of jawless vertebrates — because the fossil, armoured, jawless vertebrates known as ostracoderms (in red), are more closely related to jawed vertebrates than to hagfishes and lampreys. This implies that, during a long portion of the Early Cambrian or before, vertebrates had no mineralized tissues and could only be fossilized in exceptional

circumstances. The two new fossils (asterisks) discovered by Shu *et al.*¹ in Chengjiang, China, display such exquisitely preserved details of their soft anatomy that there is little doubt that they are vertebrates, although their phylogenetic position remains speculative. BP, before present.



Box 1: How do you recognize a vertebrate?

All living vertebrates share a relatively small number of characters found in no other group, which are assumed to have been inherited from a common vertebrate ancestor. All these characters are made of initially non-mineralized tissues (including cartilage) and are thus unlikely to be preserved in very early fossil vertebrates. The two Early

Cambrian fossils from China display some vertebrate characters directly, whereas others can be inferred. Both show a dense imprint of what is probably a cartilaginous skull, one of the main vertebrate features. The gill skeleton may be indirect evidence for a neural crest, the main vertebrate character. The fin rays (radials) in one of the

two forms and a large heart, lying behind the gills and possibly enclosed in a pericard, are also unique vertebrate characters. Chevron-shaped muscle blocks also occur in cephalochordates (the closest living relatives of vertebrates), but have a more complex zigzag shape in vertebrates, as can be clearly seen in at least one of the fossils. **P. J.**

than to any other vertebrate group, whereas the other appears as the sister-group to all other vertebrates but hagfishes (Fig. 1). The support for this result is admittedly tenuous; morphology-based data matrices for fossil vertebrate groups include a large number of question marks, because of missing data in fossils and uncertainties as to the interpretation of some of the soft-tissue characters seen in imprints. This is true for the two Cambrian fish. Nevertheless, the result makes broad sense, because it suggests that lampreys and their fossil relatives had diverged already in the Cambrian, as predicted by previous phylogenies.

Lacking fossils, palaeontologists and anatomists have often tried to imagine the earliest vertebrates. The two new fossils from China resemble some of these imaginary reconstructions, but are completely at odds with others. Were theoretical reconstructions

of the ancestral vertebrate partly accurate, or do we pay special attention to fossils that resemble the reconstructions in our minds? ■

Philippe Janvier is at the Laboratoire de Paléontologie (UMR 8569 du CNRS), Muséum National d'Histoire Naturelle, 8 rue Buffon, 75005 Paris, France.

1. Shu, D.-G. *et al.* *Nature* **402**, 42–46 (1999).
2. Young, G. C. J. *Vert. Paleontol.* **17**, 1–25 (1997).
3. Sansom, I. J., Smith, M. M. & Smith, M. P. *Nature* **379**, 628–630 (1996).
4. Aldridge, R. J., Briggs, D. E. G., Smith, M. P., Clarkson, E. N. K. & Clark, N. D. L. *Phil. Trans. R. Soc. Lond. B* **340**, 405–421 (1993).
5. Pridmore, P. A., Barwick, R. E. & Nicoll, R. S. *Lethaia* **29**, 317–328 (1997).
6. Schultze, H.-P. *Mod. Geol.* **20**, 275–285 (1996).
7. Smith, M. P., Sansom, I. J. & Repetski, J. E. *Nature* **380**, 702–704 (1996).
8. Peel, J. S. *Rapp. Grønland Geol. Undersøg.* **91**, 111–115 (1979).
9. Janvier, P. *Palaeontology* **39**, 259–287 (1996).
10. Conway Morris, S. in *Atlas of the Burgess Shale* (ed. Conway Morris, S.) 26 (Palaeontol. Assoc., London, 1982).
11. Chen, J.-Y. *et al.* *Nature* **377**, 720–722 (1995).
12. Shu, D.-G., Conway Morris, S. & Zhang, X.-L. *Nature* **385**, 865–868 (1996).

Particle physics

Maxwell's other demon

Frank Wilczek

Worrying over Maxwell's demon¹ has sharpened our understanding of the asymmetry between past and future in thermodynamics². In his lecture opening the Cavendish Laboratory in Cambridge³, Maxwell touted another demon, the Devil in the Details: "The history of science shows that even during that phase of her progress in which she devotes herself to improving the accuracy of the numerical measurements of quantities with which she has long been familiar, she is preparing the materials for the subjugation of new regions, which would have remained unknown if she had been contented with the rough methods of her early pioneers".

In rendering homage to Maxwell's other demon, physicists are again sharpening their understanding of the asymmetry between past and future, this time as it relates to the

interactions of elementary particles. The details arising from careful experiments^{4,5} and new calculations⁶ are proving surprisingly difficult to reconcile.

In the past few months, experimenters at Fermilab⁴ in the United States and at CERN⁵ in Europe announced results from their delicate, long-running, heroic experiments to measure CP violation — the asymmetry responsible for the dominance of matter over antimatter in the Universe (Box 1). To establish CP violation, particle physicists compared rare decay modes of exotic particles known as kaons. Their results are consistent with each other and with some earlier, less precise measurements⁷. They measure a critical parameter, r , describing these rare events to be $21 (\pm 4.6) \times 10^{-4}$.

The fact that r differs from zero at last puts to rest the 'superweak' model of CP vio-

lation⁸, which postulated a superweak interaction outside the Standard Model of particle physics. This is a milestone achievement, but it hardly comes as a surprise. The Standard Model beautifully accommodates the phenomenon of CP violation⁹, and it predicts quite a different, richer pattern of effects from the superweak model. Indeed, the Standard Model supplies plausible answers to two of the deepest questions this phenomenon poses: why is CP — or equivalently T — violated and why are its effects so hard to observe?

According to our current best understanding, CP violation emerges as a rather accidental by-product of the simplicity of the Standard Model. Defining a symmetry for a set of equations (that is, physical laws) means transforming the various terms in these equations, in such a way that the content of the transformed equations is precisely the same as that of the original ones. The attempt to define a symmetry in this way will fail only if one runs into a conflict — if one term in the equations requires you to transform a quantity in one way, and another term makes a different demand. It turns out that the equations of the Standard Model are so simple that they have little potential for conflict. Indeed, if there were only two families of quarks and leptons, CP and T would be automatic. In a brilliant piece of theoretical physics, Kobayashi and Maskawa⁹ understood this, and used it to argue that there ought to be a third family. At the time, only two families — the *ude* family containing up and down quarks and electrons, and the *cs μ* family containing charm and strange quarks and muons — were known. Sure enough, the third *tb τ* family containing top and bottom quarks and tau leptons duly materialized.

The fact that in the Standard Model, CP

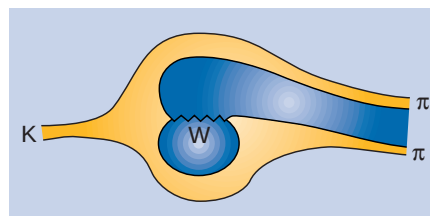


Figure 1 Many calculations in the Standard Model are neatly summarized by pictures — known as Feynman graphs — that follow the trails of virtual particles in space and time. According to Blum *et al.*⁶, the dominant contribution to the CP violation defined by r in the Standard Model comes from the Feynman graph shown above. For obvious reasons, this is called the 'eye graph'. It describes the underlying motion of quarks, antiquarks (straight lines) and a virtual W boson (zigzag) as a kaon (K) decays into two pions (π). Although eye-graph dominance comes as something of a surprise to many theorists, it could help resolve a long-standing puzzle, the so-called $\Delta I = \frac{1}{2}$ rule (don't ask!).

Box 1: Charge, Parity and Time

According to CP symmetry, the laws of physics will appear the same if you simultaneously interchange left and right (for example, reflect in a mirror) and change all particles into their antiparticles. Part of the appeal of CP symmetry is that general principles of quantum mechanics and relativity suggest that CPT — which does these transformations, and also reverses the direction of motion of all particles, or equivalently reverses the arrow of time — is a valid symmetry of nature. Assuming this, CP is then equivalent to T (ref. 15). So in testing CP symmetry, one is testing the proposition that the fundamental laws of nature do not distinguish between past and future. (Of course, this is a very different thing from saying that *the world* does not

distinguish between past and future.)

The classic laws of gravity and electromagnetism certainly do possess time-reversal or T symmetry, and the search for CP violation was fruitless for many years. Then Cronin, Fitch and colleagues¹⁶ discovered a small effect in kaon decays that is inconsistent with CP symmetry. Specifically, they found that among long-lived neutral kaons, K_L , which usually decay into three pions, about one in 1,000 particles decay into two pions ($\pi\pi$). If CP were exactly valid, that would never occur. (This is not supposed to be obvious — the argument is subtle, but quite firm.)

It has proved very difficult to improve on Cronin and Fitch. Prior to some very recent

experiments^{4,5,17}, the only firm evidence for CP violation was from kaon decays, and all of it was consistent with the 'superweak' proposal that K_L contained a small admixture of the short-lived kaon K_S , which normally decays into two pions.

The latest measurements refine Cronin and Fitch by comparing the two different possibilities for $K_L \rightarrow \pi\pi$: the pions can either be one positive and one negative, or alternatively both electrically neutral. If the superweak idea is correct, these two possibilities will occur in exactly the same ratio as they do in K_S decays. The actual ratio of ratios is defined to be $1 + 6r$. So if $r \neq 0$ the superweak proposal is wrong, and there's more to CP violation than a tainted K_L .

F. W.

(or T) violation arises only from effects of the third family explains why its effects are ordinarily so small. We usually study matter made out of just the first two families, so the only CP-violating effects occur through the small, indirect influence of virtual heavy particles from the third family. Ambitious plans are afoot to study particles containing *b* quarks directly. When such particles decay all three families are actively in play, and CP-violating effects are expected to be accessible¹⁰.

What does the Standard Model predict for *r*? This has been the object of some controversy. A major review of the subject¹¹ quotes $r = 7.7 \times 10^{-4}$, but allows for values from 4.2×10^{-4} to 13.7×10^{-4} . The reason for the spread in this prediction is revealing. Uncertainty arises not so much from any defect in fundamental theory or ignorance of fundamental parameters in the Standard Model, but from weaknesses in our ability to calculate its consequences. For example, there is no simple, reliable way to solve the equations of QCD (quantum chromodynamics) with good accuracy. In making their estimates, theorists resort to rather crude models and drastic approximations, which can come to resemble art more than science.

Into the breach have rushed the number-crunchers, creating a bombshell⁶. By solving the equations of QCD numerically, using ultrafast parallel computers, one obtains, in principle, definitive answers. A group of researchers from Columbia University and Brookhaven National Laboratory now reports that the Standard Model predicts *r* to be negative — actually, $-120 (\pm 70) \times$

10^{-4} — in dire contradiction with the experimental result. Their result reinforces earlier hints of trouble from less definitive numerical work¹². Moreover, they suggest that the difference between these calculations and previous predictions lies in their treatment of a particular process described by the 'eye graph' shown in Fig. 1.

Because the source of CP violation in the Standard Model is in principle well understood and in some sense small, experiments

in CP violation are especially sensitive to physics outside the Standard Model. Indeed, popular supersymmetry models feature so many potential new sources of CP violation that it is somewhat embarrassing that no deviation from the Standard Model has been observed (until perhaps now). Supersymmetry implies the existence of particles even heavier than the particles in the third family, although there is no evidence for them yet.

As always, extraordinary results merit extraordinary scrutiny. If the Columbia/Brookhaven result holds up, the value of *r* will join neutrino oscillations¹³ as the first indications of physics beyond the Standard Model. Already there are serious proposals for how it might be tied up with supersymmetry¹⁴.

Frank Wilczek is at the Institute for Advanced Study, School of Natural Sciences, Olden Lane, Princeton, New Jersey 08540, USA.

e-mail: wilczek@sns.ias.edu

1. Maxwell, J. C. *Theory of Heat* 328 (Longmans, Green, London, 1871).
2. Szilard, L. Z. *Phys.* **53**, 840–856 (1929).
3. Maxwell, J. C. *Scientific Papers* Vol. 2 (ed. Niven, W.) 244 (Dover, New York, 1965).
4. Alavi-Harati, A. et al. *Phys. Rev. Lett.* **83**, 22–27 (1999).
5. Debu, P. <http://www.cern.ch/NA48>
6. Blum, T. et al. <http://xxx.lanl.gov/abs/hep-lat/9908025>
7. Barr, G. et al. *Phys. Lett. B* **317**, 233–242 (1993).
8. Wolfenstein, L. *Phys. Rev. Lett.* **13**, 562–564 (1964).
9. Kobayashi, M. & Maskawa, K. *Prog. Theor. Phys.* **49**, 652–657 (1973).
10. Ciuchini, M. et al. <http://xxx.lanl.gov/abs/hep-ph/9704274>
11. Bosch, S. et al. <http://xxx.lanl.gov/abs/hep-ph/9904408>
12. Pekurovsky, D. & Kilcup, G. <http://xxx.lanl.gov/abs/hep-lat/9903025>
13. Wilczek, F. *Nature* **394**, 13–15 (1998).
14. Eyal, G., Masiero, A., Nir, Y. & Silvestrini, L. <http://xxx.lanl.gov/abs/hep-ph/9908382>
15. Weinberg, S. *The Quantum Theory of Fields I* (Cambridge Univ. Press, 1995).
16. Christenson, J., Cronin, J., Fitch, V. & Turlay, R. *Phys. Rev. Lett.* **13**, 138–140 (1964).
17. Abe, F. et al. *Phys. Rev. Lett.* **81**, 5513–5518 (1998).

Genomics

Functional links between proteins

Andrej Šali

Genome-sequencing projects have accomplished a monumental feat in generating complete lists of the proteins that make up multi-subunit assemblies and signalling pathways in various organisms. These assemblies and pathways must now be mapped, and papers by Marcotte *et al.*¹ and Enright *et al.*² (pages 83 and 86 of this issue) take a significant step in this direction. The two groups have developed computational methods that associate proteins through properties other than the similarity between their amino-acid sequences. By comparing phylogenetic (evolutionary) profiles and expression patterns, as well as by analysing domain fusions, the new methods identify proteins that are functionally linked

through a metabolic pathway, a signalling pathway or a structural complex (Fig. 1, overleaf). About half of the uncharacterized proteins in yeast — roughly a quarter of all yeast proteins — may be partially annotated by these methods¹. And, because they do not rely on direct sequence similarity, the methods can assign functions to proteins that lack detectable homologues of known function. They will find many applications in genomics, complementing experiments for the large-scale identification of protein function.

The critical information needed to construct useful models of pathways and assemblies is provided by experiment, most importantly by proteomics and structural biology. Proteomics aims to identify and

characterize complete sets of proteins and protein-protein interactions. It involves large-scale approaches such as the yeast two-hybrid system, mass spectrometry, two-dimensional gel electrophoresis and DNA-microarray hybridization³. The size and complexity of the task can be appreciated by assuming between five and 50 functional links per protein, resulting in 30,000 to 300,000 links for a single yeast cell. Although experiments have characterized about 30% of yeast proteins, they are sometimes not rapid, inexpensive or complete enough. So there is a need to assign function using computational methods.

Computational methods have traditionally assigned function by sequence similarity to a characterized protein^{4,5}. Such annotation is possible because evolution produced families of homologous proteins that share a common ancestor, and thus have similar sequences, structures and, often, functions. Protein comparisons have allowed some insight into the function of another 30% of yeast proteins⁶. However, functional assignment by homology is limited by two factors. First, it can be applied only to proteins with detectable homologues of known function. Second, it is not always clear what functional properties are shared by the matched proteins, especially for the more distant matches.

The new methods developed by Marcotte *et al.*¹ and Enright *et al.*² are not subject to these limitations because they do not depend on sequence similarity between uncharacterized proteins and proteins of known function. Instead, they group proteins that are part of the same pathway or assembly (Fig. 1) and define them as being 'functionally linked'. Marcotte *et al.* have applied three different classification schemes to the proteins in the budding-yeast genome: phylogenetic profiles⁷, domain-fusion analysis⁸ and correlated messenger RNA expression patterns⁹. The domain-fusion analysis was developed independently by Enright *et al.*, using a new clustering algorithm, and applied to three prokaryotic genomes.

Phylogenetic profiling relies on the correlated evolution of proteins^{1,7}. The evolution of two proteins is correlated when they share a phylogenetic profile, which is defined as the pattern of a protein's occurrence over a set of genomes¹⁰. The phylogenetic profile can be calculated rigorously only when several complete genomes are compared. Two proteins that share a similar phylogenetic profile are expected to be functionally linked. So, clustering of proteins based on their phylogenetic profiles can provide information about the function of an uncharacterized protein that is grouped with one or more functionally defined proteins.

The domain-fusion analysis identifies fusion proteins consisting of two non-homologous component proteins found separately in another genome^{1,2}. Such com-

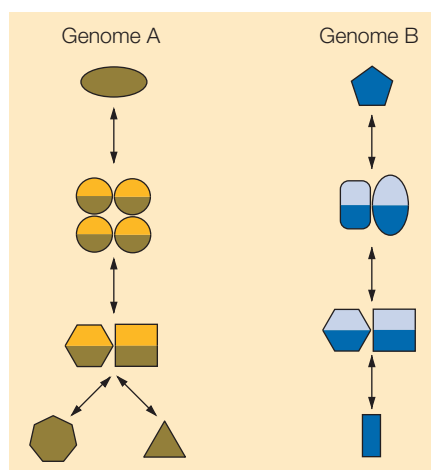


Figure 1 Functional annotation by computation. Two unrelated pathways consisting of several individual proteins and protein complexes are shown. The sequence and structural similarities of the proteins are indicated by a common shape; the functional links are indicated by a common colour. Some proteins have several links because they are part of the same pathway as well as the same assembly. Traditional annotation methods, which use sequence homology, attempt to annotate by shape. In contrast, the phylogenetic profile¹, domain-fusion analysis^{1,2} and expression correlation¹ methods developed by Marcotte *et al.*¹ and Enright *et al.*² attempt to annotate by colour.

ponent proteins are expected to interact physically with each other. An interface between two interacting component proteins is more likely to evolve when the proteins are fused in a single chain. A well-known example is fusion of the α and β subunits of tryptophan synthetase from bacteria to fungi. In some respects, the domain-fusion analysis is similar to the use of gene clusters for inferring functional links from gene proximity^{11,12}.

Marcotte *et al.* have also grouped yeast proteins by correlating their mRNA expression patterns^{1,9}. These patterns were obtained from 97 publicly available DNA chip data sets, which indicate how the expression levels of most yeast proteins change during normal growth, glucose starvation, sporulation and expression of mutant genes. The analysis is based on the expectation that proteins with correlated expression levels over the same series of conditions are functionally linked.

The new functional annotations are often broad, pinning a protein's function down to, say, 'metabolism' or 'transcription'. Even a random pair of proteins have a 50% chance of similar function at such a broad level¹. But because the annotations are generally derived from a number of linkages, they are three to eight times more informative than random links — comparable, in the best case, to experimental determination of protein-protein interactions¹. For example,

Marcotte *et al.* established new links for MSH6, a DNA-mismatch-repair protein involved in some colorectal cancers, to the PMS1 mismatch-repair family, mutations in which are also tied to human colorectal cancer, the purine-biosynthetic pathway, two RNA-modification enzymes and an uncharacterized protein family, which may now all be investigated in the light of nucleic-acid repair or modification.

How accurate are the annotations, and what percentage of proteins can they cover? These questions can be only partially addressed, because a reference set of functionally linked proteins is not easily available. Marcotte and colleagues assigned a general function to about half of the 2,557 uncharacterized proteins in yeast¹³. They estimated that up to 30% of the pairwise predictions contributing to functional assignments were false, although for the subset of predictions made by two or three of the methods together the rate of false positives decreased to 15%.

Enright *et al.* functionally linked only 215 proteins in three prokaryotic genomes by their domain-fusion analysis, but with very few estimated false positives. Their smaller rate of functional linking seems to be due to the missing links provided by the phylogenetic profiling and mRNA expression methods, which these authors did not use, a stricter definition of fusion events, and the use of fewer proteins to detect fusion. Despite false positives and coarse functional annotation, the computational methods allow experimentalists to concentrate on promising interactions. As more genome data become available, the number of predictions and accuracy of both the domain-fusion analysis and the phylogenetic profiling methods will increase.

The next step is to improve the coverage, accuracy and precision of methods for predicting protein function. This could, in theory, be done by considering three-dimensional structures, because a protein's function is determined more directly by its structure and dynamics than by its sequence. So why have structures not been used as widely as sequences in genomics? There are at least two reasons. First, three-dimensional structures are available for only a fraction of proteins. But this limitation should be reduced by structural genomics within a few years¹⁴. Structural genomics aims to determine the structures of around 10,000 carefully chosen protein domains, such that all other protein sequences can be modelled with useful accuracy¹⁵. Second, functional details that can be extracted from structure but not from sequence often depend on the details of that structure in the cellular environment, as well as on its dynamics and energetics, all of which are difficult to obtain by existing experimental and theoretical techniques.

As well as making predictions, bioinformaticians are facing the more practical challenge of making others aware of their predictions. Prediction methods need to be evaluated rigorously and made accessible over the internet¹⁶. Moreover, varied experimental data and theoretical predictions must be integrated, because no single experimental or computational approach is likely to result in accurate and complete models of protein assemblies and pathways. The latest computational methods for mapping functional links should make a big contribution to such models.

Andrej Šali is at the Laboratories of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

e-mail: sali@rockefeller.edu

- Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. *Nature* **402**, 83–86 (1999).
- Enright, A., Iliopoulos, I., Kyripides, N. C. & Ouzounis, C. A. *Nature* **402**, 86–90 (1999).
- Mendelsohn, A. R. & Brent, R. *Science* **284**, 1948–1950 (1999).
- Koonin, E. V., Tatusov, R. L. & Galperin, M. Y. *Curr. Opin. Struct. Biol.* **3**, 355–363 (1998).
- Bork, P. & Koonin, E. V. *Nature Genet.* **18**, 313–318 (1998).
- Chervitz, S. A. et al. *Nucleic Acids Res.* **27**, 74–78 (1999).
- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. *Proc. Natl Acad. Sci. USA* **96**, 4285–4288 (1999).
- Marcotte, E. M. et al. *Science* **285**, 751–753 (1999).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
- Gaasterland, T. & Ragan, M. J. *Microb. Comp. Genomics* **3**, 305–312 (1998).
- Dandekar, T. et al. *Trends Biochem. Sci.* **23**, 324–328 (1998).
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. *Proc. Natl Acad. Sci. USA* **96**, 2896–2901 (1999).
- Mewes, H. W. *Nucleic Acids Res.* **27**, 44–48 (1999).
- Burley, S. K. et al. *Nature Genet.* **23**, 151–157 (1999).
- Sánchez, R. & Šali, A. *Proc. Natl Acad. Sci. USA* **95**, 13597–13602 (1998).
- Brenner, S. E., Barken, D. & Levitt, M. *Nucleic Acids Res.* **27**, 251–253 (1999).

Geophysics

Latest spin on the core

F. A. Dahlen

Is the Earth's inner core rotating slightly faster than the mantle and crust? This is a question that has perplexed geophysicists since it was first raised five years ago. The latest study comes from Laske and Masters (page 66 of this issue¹), who have taken a new approach to the problem and conclude that the higher estimates of relative rotation rate can be ruled out.

The centre of the Earth consists of a nearly spherical ball of crystalline iron, 1,220 kilometres in radius; this solid inner core is surrounded by a fluid outer core of molten iron, whose motions are the source of the geomagnetic field. Numerical models of the geodynamo^{2,3} suggest that electromagnetic torques might cause the solid inner core to exhibit a slow differential rotation with respect to the silicate mantle and overlying crust, upon which we live. On the other hand, gravitational torques exerted by slight irregularities in the boundary topography and density of the solid inner core and mantle should keep them locked in a state of co-rotation, unless the viscosity of the inner core is low enough to allow it to deform on the requisite timescale⁴. Small but systematic temporal variations in the travel times of seismic waves passing near the centre of the Earth have been interpreted as evidence that the solid inner core is, in fact, rotating faster than the mantle and crust at a rate between 1° yr^{-1} (ref. 5) and 3° yr^{-1} (ref. 6). Subsequent seismic travel-time studies have, however, found lower differential rotation rates, ranging from $0.2^\circ\text{--}0.3^\circ \text{ yr}^{-1}$ (ref. 7) to essentially zero⁸.

Laske and Masters¹ now describe an alternative and preferable seismological pro-

cedure for measuring inner-core differential rotation, based upon temporal variations of the splitting of the Earth's large-scale free oscillations, rather than upon the travel times of short-wavelength seismic waves. These latter waves, and the whole-Earth free oscillations (which are analogous to the vibrational modes of a bell), are both generated by large earthquakes. Laske and Masters' analysis, which covers 20 years of terrestrial free-oscillation data, confidently rules out a differential rotation rate as high as 1° yr^{-1} , but is "marginally consistent" with a rate as low as $0.2^\circ\text{--}0.3^\circ \text{ yr}^{-1}$.

The principal difficulty in using seismic travel times to probe the inner core is the sensitivity of the measured times to local structural irregularities. It is like trying to measure the rotation rate of a slowly spinning, silvered globe by observing the deflection of a laser beam reflected from its surface. A beam aimed at a perfectly spherical patch will not suffer any deflection no matter how rapidly the globe is spinning; reflection from a large-scale surface irregularity will give rise to a smaller deflection than reflection from a small-scale irregularity. In the seismic problem, the waves travel through the solid inner core rather than reflect from it, and the measured observable is the travel time rather than the angular deflection.

The most persuasive example of temporal variation in seismic travel time is still the one discovered by Song and Richards⁵. Their data came from 30-year records of the arrival times of waves from earthquakes near the South Sandwich Islands at the seismic station in College, Alaska. They found a 0.3-s increase in the time difference between seismic waves that skimmed the inner core and those that passed through it. To convert this 0.01 s yr^{-1} temporal variation into their estimated rotation rate, 1° yr^{-1} , Song and Richards assumed that the South Sandwich to College path sampled only a large-scale irregularity of the inner core, namely a slight inclination of its fast axis of anisotropy.

In another study, Su et al.⁶ used a much larger travel-time data set and two different analysis procedures. But they likewise assumed that only large-scale inner-core asphericity is significant, in deducing a rotation-rate estimate of 3° yr^{-1} . Creager⁷ subsequently used a regional seismic network in Alaska to measure the present-day pattern of spatial variation in differential travel time from South Sandwich earthquakes directly.

Box 1: Free oscillations and core rotation

Every normal mode of oscillation of the Earth provides an independent large-scale image of the interior asphericity, known as a splitting function. Loosely speaking, the splitting function $f(\theta, \phi)$ of a mode is a measure of the frequency at which that mode would vibrate if the entire Earth had a spherically symmetrical structure identical to that underlying the point at colatitude θ and longitude ϕ . Only a handful of observably split oscillations 'feel' sufficiently deeply enough into the Earth to be influenced by the inner core. Sharrock and

Woodhouse⁹ looked for a slow longitudinal shift in the splitting functions of several core-sensitive modes, but their results were equivocal, because of the paucity of suitable earthquake recordings from 20 years ago.

Laske and Masters' approach¹ involved improving the 20-year comparison procedure by first using data from recent earthquakes recorded by the extensive modern global seismic network to determine accurate splitting functions $f = f^{\text{core}} + f^{\text{mantle}}$ for several core-sensitive modes, and then correcting for the

effect of aspherical mantle structure using models of f^{mantle} constrained by independent data. Regarding the mantle as stationary, the authors then sought the rotation that provides the best fit of $f^{\text{rotated core}} + f^{\text{mantle}}$ to the sparser data from past earthquakes. If the inner core is rotating rigidly, then the results from all modes and all past earthquakes should be consistent.

The results achieved with this approach led the authors to conclude that there is either no, or only very slow, differential rotation of the inner core.

F. A. D.

The observed longitudinal gradient, 0.03–0.05 s per degree, is too steep to be the result of a tilted anisotropy axis. If it is caused by a small-scale structural irregularity within the inner core, as Creager argues, then the relative rotation rate must be fairly slow: about 0.2° – 0.3° yr $^{-1}$.

Laske and Masters' method deftly sidesteps the need to resolve small-scale inner-core structure. The long-wavelength free oscillations of the Earth are, by their very nature, completely insensitive to small-scale irregularities. It is as if the spinning mirror were bathed in a sea of very long-wave radiation rather than being probed by a slender, short-wave laser beam; the resulting image of the globe would be fuzzy, with small-scale details filtered out, but any temporal variations in the filtered large-scale image could confidently be used to measure the rate of rotation. The details of Laske and Masters' approach is described in Box 1. Their conclusion is that the inner-core mean rotation rate is $0.0^{\circ} \pm 0.2^{\circ}$ yr $^{-1}$, which, as they say, "is obviously insignificantly different from zero". Creager's rate of 0.2° – 0.3° yr $^{-1}$ cannot be dismissed at the 2σ level.

So, for the moment, we have two possibilities. First, that the inner core is gravitation-

ally locked to the mantle and crust, and that the 0.01 s yr $^{-1}$ temporal variation in the differential travel time between the South Sandwich Islands and Alaska is an artefact of systematic errors caused by inhomogeneities in the seismic catalogue and changes in earthquake-location procedures. Or, second, that the inner core is rotating at 0.2° – 0.3° yr $^{-1}$ with respect to the mantle and crust, and there is not yet a long enough free-oscillation database to detect and reliably constrain such a slow rate. Only time will tell — in 2020 we will be able to compare normal-mode splitting functions obtained from recordings of the next two decades' earthquakes with those measured by Laske and Masters today. ■

F. A. Dahlen is in the Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: fad@princeton.edu

1. Laske, G. & Masters, G. *Nature* **402**, 66–69 (1999).
2. Glatzmaier, G. A. & Roberts, P. H. *Science* **274**, 1887–1891 (1996).
3. Kuang, W. & Bloxham, J. *Nature* **389**, 371–374 (1997).
4. Buffett, B. A. *Nature* **388**, 571–573 (1997).
5. Song, X. & Richards, P. G. *Nature* **382**, 221–224 (1996).
6. Su, W., Dziewonski, A. M. & Jeanloz, R. *Science* **274**, 1883–1887 (1996).
7. Creager, K. C. *Science* **278**, 1284–1288 (1997).
8. Souriau, A. *Geophys. J. Int.* **134**, F1–F5 (1998).
9. Sharrock, D. S. & Woodhouse, J. H. *Earth Planets Space* **50**, 1013–1018 (1998).

Protein folding

Linking catalysts to chemistry

Robert Freedman

For many years, a family of folding catalysts — the protein disulphide isomerases (PDIs) — has been thought to ensure the correct folding of those proteins destined for the cell surface or for secretion. Yet PDIs have never been observed doing this in intact cells. Until now. By stretching analytical and labelling techniques to their limits, Molinari and Helenius¹ report (on page 90 of this issue) that these catalysts form transient chemical links with the proteins on which they act. The authors observed PDIs covalently linked both to partly made proteins and to complete proteins that have not yet left the endoplasmic reticulum, the internal cellular compartment in which they fold to their correct conformation.

How are proteins constructed within the architecture of the cell? Electron microscopists, such as the Nobel prizewinner George Palade², showed that proteins destined for secretion from the cell pass along a complex secretory pathway. They are synthesized by ribosomes, which 'sit' on the outside of the endoplasmic reticulum. While its 'backbone' is being constructed the protein passes across the endoplasmic reticulum, so it ends up in the internal 'lumen'. (The winner of this year's Nobel Prize for Physiology

or Medicine, Günter Blobel, was honoured for his part in identifying how proteins are targeted in this way³.) The protein folds into its correct three-dimensional structure within the lumen, then proceeds to the cell surface. It does so through a series of membrane-enclosed channels and vesicles, which all fuse and bud so their cargo never has to move directly through a membrane.

Because such proteins ultimately operate outside cells (in the bloodstream or digestive tract, for example), they experience more variable and potentially destabilizing conditions than do proteins in defined and controlled intracellular environments. At a chemical level, all proteins are linear polymers, comprising a continuous backbone decorated by side chains. They fold to compact shapes owing to a complex network of weak interactions between chemical groups in the backbone and side chains. But extracellular proteins go one better — they form real covalent chemical bonds (disulphide bonds) between the side chains of cysteine residues. These cross-links are stable in the oxidizing conditions outside cells, so proteins that contain them are difficult to unfold⁴.

Disulphide bonds are formed by the PDI family in the lumen of the endoplasmic retic-

ulum. The first PDI was isolated in the early 1960s by yet another Nobel prizewinner, Christian Anfinsen. He showed that, in the test-tube, this PDI could catalyse the formation and rearrangement (isomerization) of disulphide bonds. He speculated that PDI ensures rapid formation of the disulphide bonds that allow newly made proteins to fold to the most stable structure⁵. It took nearly 25 years for this proposal to be generally accepted, but, by the late 1980s, a combination of genetic, biochemical and cell-biological evidence had established it beyond doubt⁶.

Some of this evidence showed that, by adding reactive chemicals to cells, a short chemical bridge could be formed between a PDI and the protein it was assumed to be folding. (The added chemicals were short chains with reactive groups at either end, so they could link any two proteins in close contact⁷.) The chemistry of PDI suggests that it acts by initially forming a disulphide bond between itself and the protein it is working on (its 'substrate')⁸. If so, it should be possible to detect such 'disulphide-linked' complexes. This is the challenge that Molinari and Helenius¹ decided to take on.

The technical difficulties are daunting. First is the need to recognize, in the complex mix of proteins in a cell, those that have just been synthesized. Then a specific chemical link needs to be identified between these and other proteins. A key part of the authors' strategy was the use of virus-infected cells. The Semliki Forest virus (SFV) converts cells into efficient factories for making SFV proteins. Two of these viral proteins pass through the secretory pathway — just like the cell's own secreted proteins — and ultimately form part of the SFV external membrane. By infecting cells with SFV and, some hours later, incubating them for two minutes in radioactive amino acids, Molinari and Helenius radiolabelled these viral membrane proteins as they were being made. They then followed these proteins over the next hour or more, as they made their way through the secretory pathway.

A second key aspect of the strategy was the analytical method used to determine the chemical status of the labelled SFV proteins. Molinari and Helenius took gel electrophoresis — a standard technique that separates proteins based on their size — but, crucially, separated the proteins in two stages under different conditions. In the first stage, they resolved the proteins in a thin cylinder of gel under conditions that would preserve any disulphide bonds (and inhibit any change in disulphide-bond partners). In the second stage, they separated the proteins in a slab of gel, in a direction perpendicular to the first separation, under conditions that break all disulphide bonds. Most proteins show the same mobility under both conditions, so finish up along a diagonal of the gel slab. But any proteins that were initially linked to other

proteins by disulphide bonds will lie off the diagonal, because they are bigger (less mobile) in the first dimension, then smaller (more mobile) in the second, after the disulphide bridge has been broken. The authors used a method that detects radioactivity to analyse the gels, so only the newly made SFV proteins were registered.

Putting these techniques together, Molinari and Helenius showed that, at around ten minutes after synthesis, a small fraction of the labelled SFV proteins were disulphide-linked to cellular proteins. No more than 1% of the SFV protein was in this state at any time, and no disulphide links to other proteins were found one hour later. So are the results meaningful? They are, because the authors also identified the 'partner' proteins to which the newly synthesized proteins were linked. They used a panel of antibodies to try and detect known 'molecular chaperones' and folding catalysts. The only ones that they found linked to the labelled proteins were PDI and a close relative, ERp57. The ERp57 protein works in association with two other proteins⁹, which enable it to recognize newly synthesized 'glycoproteins' (proteins with attached carbohydrate).

These results show that PDIs form transient disulphide bonds with the proteins on which they act. Evidence that the findings are not an artefact includes the fact that the labelled proteins were linked to these PDIs only, and that the links were disulphide bonds (because they were broken at the sec-

ond stage of gel separation). The results also lead to two further conclusions. First, the two PDIs show different specificities — whereas PDI links almost exclusively to one of the viral proteins, ERp57 links to both of them. And second, the PDIs act not only on complete proteins, observed ten minutes after labelling, but also on incomplete proteins detected just after labelling.

Where next? It would be interesting to obtain comparable results for the various disulphide-bonded proteins made by uninfected cells of the liver, pancreas, lymphocytes and so on. But that could be difficult. Molinari and Helenius have used a very specialized system in which a handful of well-defined proteins are made in huge amounts. Even then, the results are close to the limits of analysis, so it is not clear whether the same technique could yield results under even more challenging circumstances. But a key point has finally been established, and this will encourage those in the field to try to extend the results to other situations. ■

Robert Freedman is in the Research School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.

e-mail: r.b.freedman@uk.ac.uk

1. Molinari, M. & Helenius A. *Nature* **402**, 90–93 (1999).
2. Palade, G. *Science* **189**, 347–358 (1975).
3. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
4. Pace, C. N. *Trends Biochem. Sci.* **15**, 14–17 (1990).
5. Goldberger, R. F. *et al. J. Biol. Chem.* **238**, 628–635 (1963).
6. Freedman, R. B. *Cell* **57**, 1069–1072 (1989).
7. Roth, R. A. & Pierce, S. B. *Biochemistry* **26**, 4179–4182 (1987).
8. Freedman, R. B. *et al. Trends Biochem. Sci.* **19**, 331–336 (1994).
9. Oliver, J. D. *et al. Science* **275**, 86–88 (1997).

Galaxy formation

Clumps that survive to tell a tale

Sidney van den Bergh

How did the Milky Way system form? The present consensus¹ is that the region inside the orbit of the Sun was mainly formed by rapid collapse², whereas the outer Galactic halo assembled more slowly by the capture of many smaller bits and pieces³. Two papers published in this issue^{4,5} bear on the origin of the halo. The first by Helmi *et al.*⁴ on page 53 shows that about 8% of metal-poor halo stars near the Sun started life in a single coherent structure (perhaps a dwarf galaxy) that was disrupted during (or soon after) the Milky Way formed. The second paper by Lee *et al.*⁵ on page 55 significantly strengthens the view that the brightest globular cluster of stars in the Milky Way may actually be the remnant nucleus of another spheroidal dwarf galaxy.

The first hard evidence⁶ for the idea that the Galactic halo might have a clumpy structure was that the nearby high-velocity star Groombridge 1830 belongs to a small group that is currently passing through the Galactic

disk. More recent observations⁷ discovered a clump of stars, roughly 4.6 kpc above the Galactic plane (1 pc = 3.26 light years), which are apparently moving together through the Galactic halo. From a survey⁸ of an area in the Galactic halo known as Selected Area 57 it was concluded that the stars in this halo field show a high degree of clumping, and that "the halo field star population may derive predominantly from accretion of stellar agglomerations, most likely dwarf galaxies".

The work of Helmi *et al.*⁴ provides strong confirmation of this hypothesis. These authors have used the proper motions of stars (measured by the HIPPARCOS satellite), in conjunction with measurements of their radial velocities and parallaxes, to derive the motion in space of a sample of 88 metal-poor stars that are located within 1 kpc of the Sun. Of these stars, seven are seen to form a tight clump (Fig. 1). Numerical simulations show that the probability of this

clumping having arisen by chance is negligible. Taken at face value these results suggest that about 8% of the metal-poor Galactic population outside the solar orbit was derived from a single source (probably a satellite galaxy) that merged with the Milky Way not long after its formation. From their observed motions the stars in this group are found to have an orbital period of ~0.4 billion years (Gyr), with maximum and minimum distances from the centre of the Milky Way of ~16 kpc and ~7 kpc, respectively. During their orbit these objects reach a maximum distance of ~13 kpc from the Galactic plane (the disk of the Milky Way). The orbital characteristics of the stellar clump discovered by Helmi *et al.* are reminiscent of those for the well-studied Sagittarius dwarf spheroidal galaxy⁹, which is presently being torn apart by Galactic tidal forces.

Was the entire outer halo formed by capture of such dwarf galaxies? One may try to answer this question by recognizing that dwarf spheroidals appear to have provided an environment that was conducive to globular cluster formation for a few billion years longer than the main body of the Galactic halo. As a result the Sagittarius dwarf galaxy contains two 'young' globulars¹⁰ (Terzan 7 and Arp 2), whereas the Fornax dwarf spheroidal galaxy contains one (Fornax 4). The fact that two of the 11 known young globulars in the Milky Way and its immediate environment are found in the Sagittarius system suggests that all of the dwarf spheroidals accreted by the Galactic halo have a — highly uncertain — combined luminosity that is 11/2 times that of Sagittarius, corresponding to about 1% of the total Galactic luminosity. The extremely luminous globular cluster M54 (NGC 6715) is situated near the centre

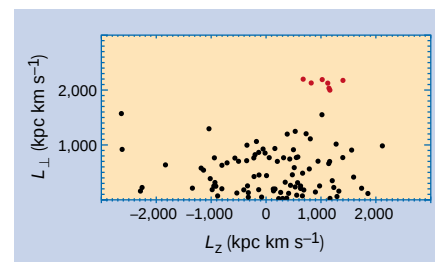


Figure 1 Clumping of metal-poor stars in the Galactic halo. From a sample of 88 stars studied by Helmi *et al.*⁴, seven are seen to form a tight clump (red points). These stars may be a fragment left over from a satellite galaxy captured by our Galaxy during (or soon after) the Milky Way formed. Together with evidence from Lee *et al.*⁵ that the brightest globular cluster in the Milky Way (ω Centauri) may be the remnant nucleus of another satellite galaxy, these results support the view that the Galactic halo was (at least in part) built up by the slow merging of such building blocks. L_z is angular momentum parallel to the Galactic plane; $L_{\perp}$ is angular momentum perpendicular to this plane.

have evolved independently from unrelated sequences. The FHA domains were first identified by a database screen centred on a conserved region in members of the Fork-head-type transcription factor family⁶. Found mainly in protein kinases, phosphatases and transcription factors, FHA domains consist of roughly 75 amino acids divided into three or four highly conserved sequence blocks. They were identified in proteins from yeast, plants, mammals and bacteria, and in many uncharacterized open reading frames.

But the function of FHA domains remained a mystery. A year before these domains were identified as distinct signalling modules, John Walker's laboratory identified a 239-amino-acid stretch within the kinase-associated protein phosphatase (KAPP) of the tiny weed *Arabidopsis thaliana* that bound a serine/threonine-receptor-like kinase, RLK5. This interaction depended on phosphorylation of RLK5 on serine or threonine residues⁷. Sequence comparisons revealed an FHA domain in the middle of this kinase-interacting domain, and earlier this year Walker and colleagues reported³ that this FHA domain is critical for the phosphorylation-dependent binding of KAPP to several protein kinases.

A second clue that FHA domains might be phosphoserine/threonine binding modules emerged from studies by David Stern and colleagues. They were searching for proteins that bind to Rad53, a protein kinase involved in controlling cell-cycle checkpoints in the yeast *Saccharomyces cerevisiae*⁸. Rad53 contains two FHA domains flanking a central protein kinase domain, and Stern's group mapped the interaction of Rad53 with another DNA-damage-control protein, Rad9, to the second of these FHA domains. Like the interactions between KAPP and RLK5, the Rad53–Rad9 coupling was specific for the phosphorylated form of Rad9.

Durocher *et al.*¹ now report that, in fact, both of Rad53's FHA domains can bind to Rad9. They show the first direct interaction between the amino-terminal FHA domain of Rad53 and short peptides containing phosphothreonine. Then, using immobilized peptides, they demonstrate that Rad53's carboxy-terminal FHA domain, along with the FHA domains from proteins in *Arabidopsis*, humans, yeast and *Mycobacterium tuberculosis*, also binds sequences that contain phosphothreonine. Although the exact motifs that FHA domains recognize — and the serine/threonine kinases that generate these motifs — are not known, all the data point to the FHA domains as 'SH2-domain equivalents' in the world of phosphoserine/threonine signalling.

But there are other phosphoprotein-binding domains. Some WW domains, for example, bind to phosphoserine. They consist of around 35–40 amino acids folded into a

three-stranded β -sheet, and bind to a variety of proline-rich proteins. One such protein, Pin1, regulates entry to and exit from mitosis, presumably by isomerizing phosphoserine–proline bonds⁹. Earlier this year, Lu *et al.*² showed that the WW domains from Pin1, the yeast Pin1 homologue Ess1 and the ubiquitin ligase NEDD4, act independently as phosphoserine/threonine-binding modules.

The most highly conserved phosphoserine/threonine-binding proteins identified to date are members of the 14-3-3 protein family^{4,5}. In contrast to modular signalling elements (such as FHA and WW domains), which are usually interspersed with other domains in signalling proteins, 14-3-3 molecules are themselves functional dimeric proteins. There are seven different 14-3-3 isotypes in mammalian cells, and at least ten in *Arabidopsis*, allowing them to act in tissue- and organelle-specific ways. When 14-3-3 proteins bind their phosphoprotein prey, they are thought to restrict the bound proteins to the cytosol, either by preventing them from entering the nucleus or by speeding their passage out of it.

Other families of phosphoprotein-binding domains probably remain to be discovered. Those that have been identified so far are unrelated in primary sequence and three-dimensional structure, so must have evolved independently. In some cases — such as phosphotyrosine-binding and WW domains — only one sub-group of the family has phosphopeptide-binding specificity. Perhaps sub-groups of other, previously defined modular domains have evolved phospho-amino-acid specificity. For example, there is evidence that WD40 repeats within F-box proteins mediate the phosphoserine/threonine-dependent binding and ubiquitin-mediated degradation of cyclin-dependent kinase inhibitors^{10,11}. The challenge for the future will be to identify the function of each phosphoserine/threonine-binding protein in specific signalling cascades. ■

Michael B. Yaffe and Lewis C. Cantley are in the Division of Signal Transduction, Harvard Medical School/Beth Israel Deaconess Medical Center, Harvard Institutes of Medicine, 330 Brookline Avenue, Boston, Massachusetts 02215, USA. e-mails: myaffe@caregroup.harvard.edu cantley@helix.mgh.harvard.edu

1. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. *Mol. Cell* **4**, 387–394 (1999).
2. Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. *Science* **283**, 1325–1328 (1999).
3. Li, J., Smith, G. P. & Walker, J. C. *Proc. Natl Acad. Sci. USA* **96**, 7821–7826 (1999).
4. Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. *Cell* **84**, 889–897 (1996).
5. Yaffe, M. B. *et al. Cell* **91**, 961–971 (1997).
6. Hofmann, K. & Bucher, P. *Trends Biochem. Sci.* **20**, 347–349 (1995).
7. Stone, J. M., Collinge, M. A., Smith, R. D., Horn, M. A. & Walker, J. C. *Science* **266**, 793–795 (1994).
8. Sun, Z., Hsiao, J., Fay, D. S. & Stern, D. F. *Science* **281**, 272–274 (1998).
9. Yaffe, M. B. *et al. Science* **278**, 1957–1960 (1997).
10. Skowrya, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. *Cell* **91**, 209–219 (1997).
11. Henchoz, S. *et al. Genes Dev.* **11**, 3046–3060 (1997).

Daedalus

Chain in miniature

Microtechnology, even nanotechnology, are popular buzzwords these days. Some of the schemes mustering under these banners may even be feasible. In particular, tiny machines, mass-produced from photo-etched silicon and operating piezoelectrically at kilocycle or megacycle rates under computer control, are already at a pilot stage. Daedalus reckons they are needed in the textile trade.

Weaving is a macrotechnological triumph, but only with continuous fibre. It cannot imitate that mediaeval masterpiece, chain mail. This armoured cloth was made of innumerable interlinked loops of metal wire. Each loop had to be individually threaded into place and closed into a circle by thermal forging. The metal links were hard and rigid; but the mail itself was wonderfully flexible. A silicon micro-machine could replicate this structure on the ten- or hundred-micron scale, as rapidly as a conventional loom can weave. It could be fed with normal polymeric fibre, to be chopped into short lengths, bent and melt-welded into linked loops. Fine-gauge tube, slit into rings which are then sliced open, might be even better. After threading, each ring would naturally close again under its own elasticity, ready for its two ends to be melted together. A silicon loom would have thousands of such threading and closing units in a line, each receiving its own feedstock, and linking it into the growing web.

'Micro-chaincloth' will transform the world of fashion. For a start, it will not be limited to a 'warp and weft' structure. Hexagonal or even denser weaves will be feasible, as will three dimensional ones, with several layers of links coupled in thickness. And like chain mail itself, the new cloth will be amazingly tough. It will distribute any load through innumerable linkages, each always capable of a little extra yielding. Yet for all that, micro-chaincloth will be magically soft. With no stiffness above the link scale, it will have a flattering, limpid, almost liquid drape. By the same token, it will be utterly snag-resistant. Having no continuous fibre, it cannot come unravelling; local damage cannot spread. In particular, it will never fray. The edges of a garment will no longer need to be folded over and stitched firm. The widespread worry about visible panty-lines will vanish for ever.

David Jones

The further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special Nature offer: m.curtis@nature.com

130th anniversary



Hildebrand in Germany, and Darwin in England, have investigated the very important part played by insects in the fertilisation of the pistil of one individual by the stamens of another individual of the same species. It is now generally admitted by botanists that cross-fertilisation is the rule rather than the exception. ... For this cross-fertilisation to take place, however, some



foreign agency like that of insects is evidently necessary ... How then is fertilisation accomplished in those plants which flower habitually in the winter, when the number of insects that can assist in it is at all events very small? ... During the winter of 1868-69, I had the opportunity of making some observations on this class of

plants; the result being that I found that, as a general rule, fertilisation, or at all events the discharge of pollen by the anthers, takes place in the bud before the flower is opened, thus ensuring *self-fertilisation* under the most favourable circumstances, with complete protection from the weather, assisted, no doubt, by that rise in temperature which is known to take place in certain plants at the time of flowering. The dissection of a flower of *Labium album* (Fig. A) gathered the last week in December, showed the stamens completely curved down and brought almost into contact with the bifid stigma, the pollen being at that time freely discharged from the anthers. A more complete contrivance for self-fertilisation than is here presented would be impossible.

THE TWIN RECORDS OF CREATION;
or, *Geology and Genesis, their Perfect Harmony and Wonderful Concord.*
By GEORGE W. VICTOR LE VAUX. Numerous Illustrations. 2s. cloth.
"The author combines an unbounded admiration of science with an unbounded admiration of the Written Record. The two together are balanced in a steady; and the consequence is, that difficulties, which to others less evenly poised would be serious, find immediate solution in the happiest hands."—*London Review*.

SCIENCE ELUCIDATIVE OF SCRIPTURE, and NOT ANTAGONISTIC TO IT; a Series of Essays on Alleged Discrepancies; The Theory of the Geological and Figure of the Earth; the Moon's Cosmogony; Miracles in Genesis; the Miracles of Joshua; The Supernaturally Impossible, &c. &c. By Professor J. B. YOUNG, Author of "A Course of Elementary Mathematics," &c. &c. Price 10s. 6d. cloth.
"Distinguished by the true spirit of scientific inquiry, by great knowledge, by keen logical ability, and by a style peculiarly clear, easy, and energetic."—*Newcomenist*.
London: LOCKWOOD & CO. 5, Bedford's Hall Court, E.C.

The "Female Physicians" question, thanks to Professor Masson, has made a great stride during the past week. Ladies are to be admitted to study Medicine at Edinburgh University. Imagine the feelings of the non-contents when Professor Masson, in a final outburst, described their argumentation as "rampageous mysticism, dashed with drivel from Anacreon!"

A selection of words and images from the first issue of *Nature*, published on 4 November 1869.

BRAGG'S VEGETABLE CHARCOAL

or PURE CARBON PISCUTIS, a nutritious, pleasant, and healthful diet, which has produced great benefit and positive relief to thousands of sufferers from indigestion, bile, acidity, foul breath, dyspepsia, heartburn, worms, &c. There is medical testimony to its beneficial effect in these complaints.—Sold in tins, 1s., 2s., 4s., and 8s. each, by all chemists, and by the manufacturer, J. L. BRAGG, 14 (near 2), Wigmore Street, Cavendish Square.

An eclipse of the sun, so beautiful and yet so terrible to the mass of mankind, is of especial value to the astronomer, because at such times the dark body of the moon, far outside our atmosphere, cuts off the sun's light from it, and round the place occupied by the moon and moon-eclipsed sun there is therefore none of the glare which we usually see — a glare caused by the reflection of the sun's light by our atmosphere. If, then, there were anything surrounding the sun ordinarily hidden from us by this glare, we ought to see it during eclipses. In point of fact, strange things are seen. There is a strange



halo of pearly light visible, called the corona ... What, then is the evidence furnished by the American observers on the nature of the corona? It is *bizarre* and puzzling to the last degree! The most definite statement on the subject is, that it is nothing more nor less than a *permanent solar aurora!* ... Formerly, a favourite argument has been that because the light of the corona is

polarised; therefore it is solar. The American observers state that the light is *not* polarised — a conclusion, as M. Faye has well put it, "very embarrassing for Science." Further, — stranger still, if possible, — it is stated that another line of inquiry goes to show that, after all, Halley may be right, and that the corona may really be due to a lunar atmosphere. ... Certainly, never before was an eclipsed sun so thoroughly tortured with all the instruments of Science. Several hundred photographs were taken, with a perfection of finish which may be gathered from the accompanying reproduction of one of them.

The public anxiety about the fate of our great explorer, Dr. Livingstone, has been anything but allayed by the recent telegrams from Bombay and Zanzibar, wanting, as they seem to do at present, the stamp of the approval of Sir R. Murchison. The Bombay mail is now hourly expected; and, by the opening meeting of the Royal Geographical Society, Sir Roderick will be in possession of all the data on which to form a complete estimate of the recent intelligence, and will communicate the results. In the meantime, we wait and hope; Livingstone is not the man to do his work hastily or incompletely, or to return leaving anything unexplored.

Nov. 4, 1869]

NATURE

37

FOR CHRISTMAS PRESENTS, ETC.

NEW & BEAUTIFUL SLIDES & EFFECTS FOR THE MAGIC LANTERN & DISSOLVING VIEWS,

ALSO THEIR NEW CHEAP STUDENT MICROSCOPE, THE BEST CHEAP MICROSCOPE MADE.

GOULD AND PORTER (Late Cary).

OPTICIANS AND MATHEMATICAL INSTRUMENT MAKERS

To the Admiralty, Royal London Ophthalmic Hospital, Royal Geographical Society, Royal Military College, Sandhurst, Trinity House, Christ's Hospital, King's College, &c. &c.

151, STRAND, LONDON. ESTABLISHED UPWARDS OF A CENTURY.

Trans-gender induction of hair follicles

Human follicle cells can be induced to grow in an incompatible host of the other sex.

Adult hair growth and the cyclic regeneration of hair follicles are driven by finely regulated interactions between dermal and epidermal tissue at the base of the follicle^{1,2}. In rodents, isolated hair-follicle dermal cells that are reintroduced *in vivo* can induce new follicles^{3,4}, but this has never been achieved in humans. Despite the great advances in human tissue and organ transplantation, rejection processes severely restrict the transfer of cells between immunologically incompatible individuals, although the hair follicle is considered to be one a group of organs that display 'immune privilege'⁵. Here we show that transplantation of human follicle dermal sheath tissue between incompatible individuals of different sex can induce the formation of new hair follicles.

To test the inductive and immunoreactive properties of human hair-follicle dermis, we microdissected dermal sheath tissue from the base of scalp skin follicles of one of us (C.A.B.J.), a male. We implanted this tissue into shallow skin wounds on the inner forearm of another one of us (A.J.R.), a genetically unrelated and immunologically incompatible female recipient (Fig. 1a–f). Five months later, the same female received grafts of follicle dermal sheath and dermal papilla from the same male donor and, in a third experiment, dermal sheath from a second unrelated male donor.

All the wound sites healed rapidly and lacked any overt inflammatory reaction. Remarkably, each of the sites of dermal-sheath implantation produced new follicles and fibres 3 to 5 weeks after the graft (Fig. 1g). Unlike the tiny, unpigmented vellus hairs of the arm, the newly induced hairs were larger and thicker, mostly pigmented, and grew in variable directions. None of the new follicles showed evidence of rejection when biopsied between 41 and 77 days after the graft. Histology confirmed that the new follicles were morphologically normal, with oval dermal papillae overlaid by a pigmented epidermal matrix at their base. We observed a light inflammatory infiltrate around the upper regions of one specimen.

To demonstrate that the follicles were induced by the implanted dermal tissue, we analysed DNA from our histological specimens. We used laser capture microdissection⁶ to isolate dermal papilla cells from one of the follicles (Fig. 1h) and performed polymerase chain reaction (PCR)-based gender determination and restriction analysis of simultaneously amplified *ZFX* and *ZFY* sequences⁷. This confirmed that the DNA extracted from the papilla cells had an

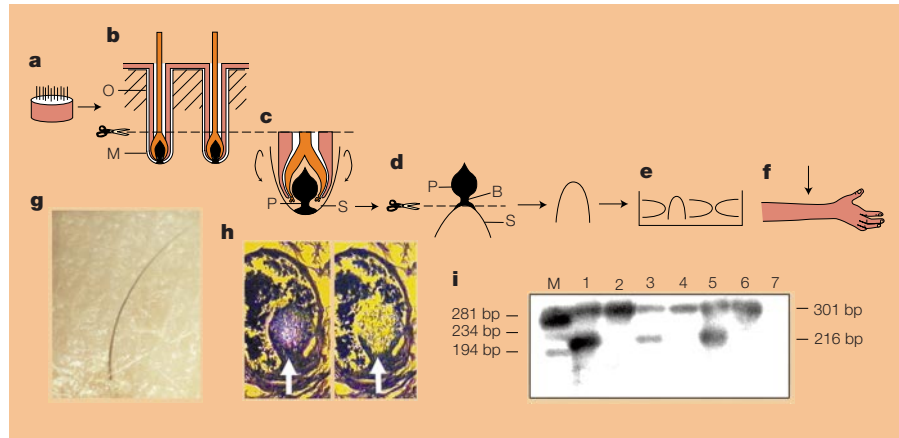


Figure 1 Methods and analysis of follicle induction. **a–d**, Punch biopsies were taken from male scalp skin (**a**) and the hair-follicle end bulbs amputated (**b**), everted (**c**) and the exposed epidermal components scraped away and discarded (**d**). **e**, The dermal sheath and dermal papillae were separated, cleaned and pooled separately in culture medium. **f**, Up to 11 pieces of dermal sheath were combined and implanted into small, shallow wounds on the inner forearm. **g**, Induced fibres were distinct and often pigmented. Laser capture microdissection (Pixi Cell, Arcturus, California) was used to isolate cells from histology slides without coverslips. **h**, From induced follicles, 200 to 400 dermal papilla cells were removed (arrows) with six 'punches', and a similar number of cells were dissected from controls. After DNA extraction, we performed PCR-based gender determination by restriction analysis of simultaneously amplified *ZFX* and *ZFY* sequences, as described⁷. Digested PCR products were Southern blotted and hybridized with a probe consisting of the same PCR product for visualization. **i**, Southern-blot analysis of the restricted PCR products shows two products with 301 and 216 base pairs (bp), indicative of the X and Y chromosomes, respectively, in DNA from male donor blood (lane 1), from the laser-capture microdissected cells of the induced follicle (lane 3), and from laser-captured skin cells from an unrelated male (lane 5). A band at 301 bp (indicating only the X chromosome) is visible in DNA from female host blood (lane 2), from laser-capture microdissected cells of the female host adjacent to the induced follicle in (lane 4), and from laser-captured skin cells from an unrelated female (lane 6). Lane 7 is a DNA-free control, and lane M contains molecular weight markers digested with *Hae*III. M, epidermal matrix; I, inner root sheath; O, outer root sheath; P, dermal papilla; B, basal stalk; S, dermal sheath. This work was approved by the Newcastle Health Authority Joint Ethics Committee, the University of Newcastle upon Tyne.

X and Y complement (and was therefore male), whereas cells taken from other regions of the same follicle and epidermis in the immediate vicinity had only X chromosomes and so were female (Fig. 1i).

We have shown that just a few hundred cells of follicle dermal-sheath tissue from the scalp of an adult human male was sufficient to form new dermal papillae and induce new hair follicles in the skin of a genetically unrelated female. Moreover, these follicles survived prolonged periods without being rejected. However, the failure of the dermal papillae to induce follicles is surprising because the papilla has previously been shown to be highly inductive⁴.

The formation of follicles requires interaction between the host epidermal and donor dermal cells, which can occur only if the male tissue is tolerated. Our donor and host were different with regard to blood group and major histocompatibility complex (MHC), but there was no strong evidence of graft rejection, even after repeated grafting, when a rapid second-set rejection might have been expected.

Grafts are more likely to be tolerated if there is no coincidental transfer of passenger leukocytes, which express the donor's

class II MHC molecules, Langerhans cells or blood-vessel endothelial cells, all of which can stimulate the host's alloreactive response^{8,9}. We washed the tissue repeatedly to remove adherent cells and it was unlikely to have included blood vessels. Small-scale tissue damage, which can also trigger the host's immune response¹⁰, was also negligible in our minimally invasive procedure.

But the main reason why the graft was not rejected probably lies with the tissue itself. Immunohistochemical investigations have previously indicated that the lowermost tissues of rat and human hair follicles have special immune status¹¹. Our findings support this idea and further indicate that the follicle-end bulb dermal sheath is among a restricted group of immunoprivileged tissues that can be transplanted to foreign sites without being rejected⁵.

Our results show that follicle dermal cells from a human adult can initiate epithelial–mesenchymal interactions and create new follicles without being rejected. This mini-organ morphogenesis shows how adult cells with inductive properties might be used in tissue and organ engineering and, more immediately, might be used in new treatments for hair loss.

Amanda J. Reynolds*, Clifford Lawrence†, Peter B. Cserhalmi-Friedman‡, Angela M. Christiano‡, Colin A. B. Jahoda*

*Department of Biological Sciences, Durham University, South Road, Durham DH1 3LE, UK
e-mail: c.a.b.jahoda@durham.ac.uk

†Department of Dermatology, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK

‡Departments of Dermatology, and Genetics and Development, Columbia University, New York, New York 10032, USA

1. Hardy, M. H. *Trends Genet.* **8**, 55–61 (1992).
2. Oliver, R. F. J. *Soc. Cosmet. Chem.* **22**, 741–755 (1971).
3. Reynolds, A. J. & Jahoda, C. A. B. *Development* **115**, 587–593 (1992).
4. Jahoda, C. A. B., Reynolds, A. J. & Oliver, R. F. J. *Invest. Dermatol.* **101**, 584–590 (1993).
5. Streilein, J. W. *Science* **270**, 1158–1159 (1995).
6. Simone, N. L. *et al. Trends Genet.* **14**, 272–276 (1998).
7. Chong, S. S. *et al. Hum. Mol. Genet.* **2**, 1187–1191 (1993).
8. Fabre, J. W., Milton, A. D., Spencer, S., Settaf, A. & Houssin, D. *Transplant. Proc.* **19**, 45–49 (1987).
9. Halloran, P. F., Broski, A. P., Batiuk, T. D. & Madrenas, J. *Transplant. Immunol.* **1**, 3–27 (1993).
10. Ibrahim, M. A. A., Chain, B. M. & Katz, D. R. *Immunol. Today* **16**, 181–186 (1995).
11. Gibson, W. T., Westgate, G. E. & Craggs, R. I. *Ann. NY Acad. Sci.* **642**, 291–300 (1991).

Conservation biology

Restoration of an inbred adder population

The negative effects of inbreeding on population size are well documented in captive animals¹, but there is surprisingly little evidence that genetic factors cause a decline in wild populations^{2,3}, apart from a reported correlation of low levels of genetic variability with a high incidence of malformed or stillborn offspring⁴. From the point of view of conservation strategies, it is not only the effect of genetic factors on population decline that needs to be considered, but also whether introducing novel genes can prevent or reverse such a decline. Here we show that the introduction of new genes into a severely inbred and isolated population of adders (*Vipera berus*) halted its precipitous decline towards extinction and expanded the population dramatically.

We studied an isolated population of adders at Smygehuk, about 50 km south of Lund in Sweden. The area is bordered to the south by the Baltic Sea, to the north by arable land, to the west by a village and to the east by a harbour. The adders are confined to a coastal strip of grassy meadow 1 km long and 50–200 m wide. They have been isolated from other adder populations for at least a century, and the nearest known population is 20 km to the north across agricultural fields unsuitable for adders⁴.

The population declined dramatically around 35 years ago⁴ and has since suffered from severe inbreeding depression, with a high proportion of deformed or stillborn offspring and very low genetic variability⁴. Since 1981, all adders captured in the study area have been marked by clipping of the ventral scales. The small size and open habitat of the study area means that all adult males can be captured during spring (late March to early May) each year, when they bask in the open, enabling us to count the new recruits (recently matured males).

Our estimates of population number are based on adult males only, as these can be most reliably censused. Female adders in

our study area have an approximately biennial reproductive cycle. Non-reproductive females emerge from hibernation up to four weeks later than reproductive females and exhibit cryptic behaviour, making them much harder to find in spring than males and reproductive females. In any given year, therefore, we are unable to find all non-reproductive females, so counting male snakes provides the most robust indicator of population dynamics.

The population size at Smygehuk has fallen each year since 1983 (the decline from 1981 to 1995 is highly significant: $r = -0.83$, $P = 0.0001$, d.f. = 14; Fig. 1a), mainly because there are fewer recruits (correlation between the total number of males captured each year and the number of recruiting (unmarked) males: $r = 0.79$, $P = 0.0001$, d.f. = 17). The population was lowest in 1995, when only four males were caught.

In spring 1992, we captured 20 adult male adders from large and genetically variable populations north of Smygehuk and released them into the Smygehuk population. They remained there for four mating seasons, and the eight surviving snakes were captured and released back into their natal populations in 1995. The introduced males are not included in our data on population demography (Fig. 1a).

The introduced males settled in rapidly, and mated with all the reproductive female adders in Smygehuk from their first season⁴. Because male adders mature at about four years of age⁵, we did not expect to see any effect on the number of adult males until 1996, which is what occurred. From 1996 to 1999, there were dramatic increases in the number of recently matured males recruiting to the population ($r = 0.96$, $P = 0.04$, d.f. = 3) and the total number of males captured ($r = 0.98$, $P = 0.02$, d.f. = 3), which peaked at 32 in 1999, the most recorded over the 19-year study (Fig. 1a).

The genetic variability within the population also increased rapidly from 1996 to 1999. We isolated DNA from whole blood and analysed polymorphism in the MHC class I by restriction-fragment length polymorphism⁶. Before the new males were introduced, the population had extremely low genetic variability (Fig. 1b). The recently recruited males exhibited much more genetic variability (Fig. 1b), confirming that most of them were sired by the introduced males. The proportion of stillborn offspring also fell suddenly⁴, indicating that the rapid increase in recruitment was due to increased survival of juvenile adders.

The isolated nature of the population precludes immigration, so the only plausible alternative would require an increase in the number of litters produced. However, the number of females reproducing each year was slightly lower during the recovery phase than during the earlier decline (mean, 8.2 litters per year from 1981 to 1991 compared with 4.8 litters per year from 1992 to 1995).

Our population data from 1983 to 1995 indicate that the Smygehuk adders were on the brink of extinction, with falling numbers and negligible recruitment (Fig. 1a). Introducing new genes from a different population enabled the adders to make a dramatic recovery. This result encourages genetic approaches to conservation and

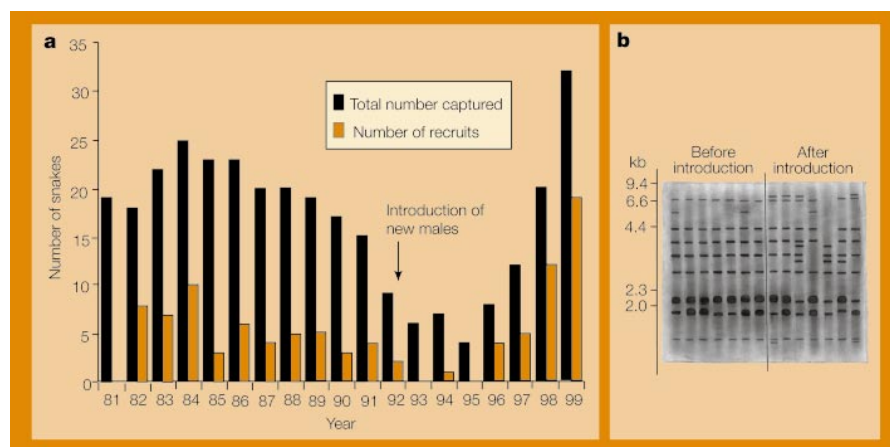


Figure 1 Introducing new males increases the genetic diversity and enables the adder population to recover. **a**, Total number and number of recruited male adders captured in Smygehuk from 1981 to 1999. **b**, Southern-blot analysis of major histocompatibility complex (MHC) class I genes in seven males sampled before the introduction of new males (left) and in seven recruited males sampled in 1999 (right).

supports the importance of preserving genetic variability as a way of increasing the viability of wild populations.

Thomas Madsen*†, Richard Shine*, Mats Olsson‡, Håkan Wittzell§

*School of Biological Sciences A08,

University of Sydney,

Sydney, New South Wales 2006, Australia

†Department of Animal Ecology and §Department

of Theoretical Ecology, Lund University,

Lund 223 62, Sweden

‡Department of Zoology, Animal Ecology,

University of Gothenburg, Box 463,

Gothenburg 405 30, Sweden

1. Ralls, K. & Ballou, J. *Trends Ecol. Evol.* **1**, 19–22 (1986).

2. Saccheri, I. *et al. Nature* **392**, 491–494 (1998).

3. Westemeier, R. L. *et al. Science* **282**, 1695–1698 (1998).

4. Madsen, T., Stille, B. & Shine, R. *Biol. Conserv.* **75**, 113–118 (1996).

5. Madsen, T. & Shine, R. *J. Anim. Ecol.* **63**, 561–568 (1994).

6. Wittzell, H., von Schantz, T., Zorob, R. & Auffray, C.

Immunogenetics **39**, 395–403 (1994).

Olfaction

The world smells different to each nostril

The flow of air is greater into one nostril than into the other because there is a slight turbinate swelling in one^{1–3}. The nostril that takes in more air switches from the left to the right one and back again every few hours^{4–6}, but the effect of this switching on the sense of smell has been unclear^{7,8}. Here we show that this difference in airflow between the nostrils causes each nostril to be optimally sensitized to different odorants, so that each nostril conveys a slightly different olfactory image to the brain.

The slight swelling that obstructs each nostril (Fig. 1a) causes odorants to be drawn into the nostrils at different rates. But for an odorant to act on the olfactory receptors, it must first cross the olfactory mucosa. Different odorants sorb to and cross the mucosa at different rates⁹. In the bullfrog, for example, a specific odorant's sorption rate interacts with the rate of airflow across the mucosa to produce varying amplitudes of response in the olfactory nerve¹⁰. A high-sorption odorant induces a smaller response when airflow is low and a larger one when it increases. In contrast, a low-sorption odorant induces a smaller response at a high airflow rate and a larger response when there is less airflow (Fig. 1b).

This occurs because, when a high-sorption odorant has a low airflow rate, the odorant molecules sorb to the mucosa before moving very far along it. Only a small portion of the epithelium is involved in the response, which is small. When the same odorant flows at a high airflow rate, it spreads across a larger mucosal area before sorbing, so the response is larger. When a low-sorption odorant flows quickly, it

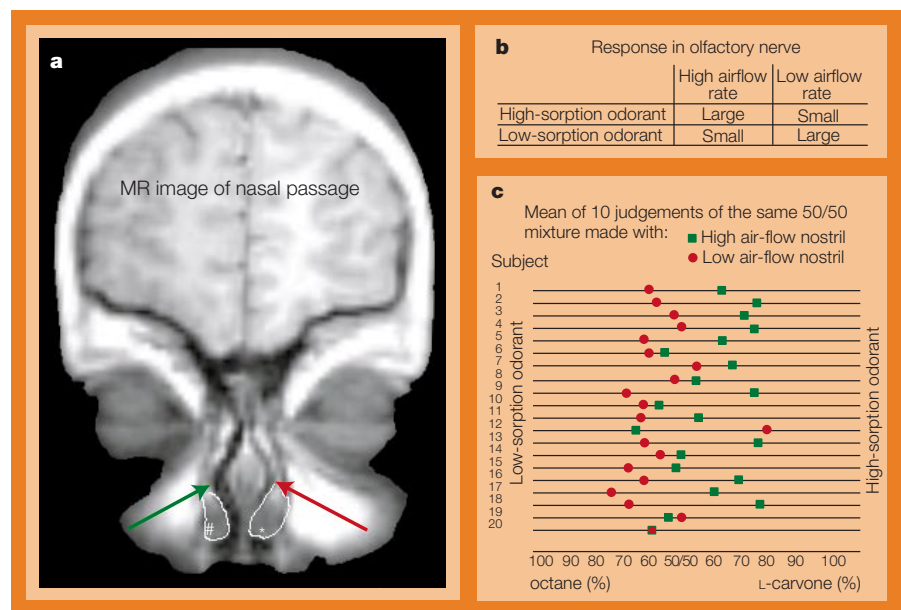


Figure 1 Different nostrils convey different olfactory information to the brain. **a**, Magnetic resonance image of the nasal passage, which appears dark. The swollen (*) and relaxed (#) turbinates, outlined in white, result in an occluded right nostril (red arrow) and a clearer left nostril (green arrow). **b**, The interaction between airflow rate and odorant sorption, which brings about a response in the olfactory nerve¹⁰. **c**, On each of ten trials, subjects were asked to smell an identical mixture of 50% octane and 50% L-carvone using either the left or right nostril. They were then given each individual odorant component to smell separately and judged the composition of the mixture by marking the line (experimental sequences were randomized and counterbalanced). Using the high-flow-rate nostril (green), the average judgement was that the mixture consisted of 55% L-carvone and 45% octane. Using the low-flow-rate nostril (red), the judgement was that it consisted of 61% octane and 39% L-carvone ($t(19) = 3.74$, $P = 0.001$). For the 20 subjects, there was no significant group difference in airflow rate between the left and right nostrils, but there was a significant group difference between the high-flow-rate nostril and the low-flow-rate nostril (high mean = 51 l min⁻¹, low mean = 31 l min⁻¹, $t(19) = 5.6$, $P < 0.0001$).

moves past the mucosa without sorbing so the epithelial response is small. When the same low-sorption-odorant flows slowly, it has time to sorb across the mucosa and the response is larger¹⁰.

We therefore investigated whether the nostril with the higher airflow in humans is the more sensitive to high-sorption odorants and the nostril with lower airflow more sensitive to low-sorption odorants. We used an olfactometer to produce an equally proportioned mixture of the high-sorption odorant L-carvone and the low-sorption odorant octane. The mixture was always the same but subjects were told that it was slightly different for every trial.

Subjects sampled the mixture by sniffing with one nostril (the other nostril was occluded) and made a judgement about the relative proportion of the two components in the mixture (for example, 55% octane and 45% L-carvone). The task was repeated for the second nostril and the judgements compared. The rate of airflow for each sniff was measured by anterior rhinomenometry. We found that 17 of 20 subjects (binomial, $P = 0.001$) thought the mixture contained more octane when they used the low-airflow nostril, and more L-carvone when they used the high-airflow nostril (Fig. 1c).

The nostril with the higher airflow reverses periodically^{4–6}, so we tested eight subjects after the nostril with greater airflow had switched, and found that the perception

of the same mixture reversed in seven of the eight subjects (binomial, $P = 0.035$). Odorant perception was therefore dependent on airflow rate, not on whether the odorant was smelled by the left or right nostril.

The different airflow between the nostrils results in a disparity of olfactory perception. Providing the olfactory system with two disparate images of the olfactory world with each sniff in this way may improve olfactory acuity by expanding the range of odorants that are within optimal sensitivity in a given sniff.

Noam Sobel*, Rehan M. Khan†, Amnon Saltman‡, Edith V. Sullivan*§, John D. E. Gabrieli*†

*Program in Neuroscience and Departments of

†Psychology and §Psychiatry and Behavioral

Science, Stanford University,

Stanford, California 94305, USA

‡Ministry of Environmental Protection,

Ramle 72100, Israel

e-mail: nsobel@leland.stanford.edu

1. Kayser, R. *Arch. Laryngol.* **3**, 101–120 (1895).

2. Principato, J. J. & Ozenberger, J. M. *Arch. Otolaryngol.* **9**, 71–77 (1970).

3. Hasegawa, M. & Kern, E. B. *Mayo Clin. Proc.* **52**, 28–34 (1977).

4. Mirza, N., Kroger, H. & Doty, R. L. *Laryngoscope* **107**, 62–66 (1997).

5. Gilbert, A. N. & Rosenwasser, A. M. *Acta Otolaryngol.* **104**, 180–186 (1987).

6. Bojsen-Muller, F. & Fahrenkrug, J. *J. Anat.* **110**, 25–27 (1971).

7. Eccles, R., Jawad, M. S. & Morris, S. *Acta Otolaryngol.* **108**, 268–273 (1989).

8. Frye, R. E. *Handbook of Olfaction and Gustation* (ed. Doty, R. L.) 471–491 (Dekker, New York, 1995).

9. Mozell, M. M. & Jagodowicz, M. *Science* **181**, 1247–1249 (1973).

10. Mozell, M. M. *et al. Chem. Sens.* **16**, 631–649 (1991).

Vision

Trichromatic vision in prosimians

Trichromatic vision in primates is achieved by three genes encoding variants of the photopigment opsin that respond individually to short, medium or long wavelengths of light. It is believed to have originated in simians because so far prosimians (a more primitive group that includes lemurs and lorises) have been found to have only monochromatic or dichromatic vision^{1–3}. But our analysis of the X-chromosome-linked opsin gene in 20 representative prosimian species provides evidence for trichromacy in ancestral and extant prosimians, indicating that it may have originated much earlier than is commonly believed.

Trichromacy in humans, apes, Old World monkeys and howler monkeys⁴ (a genus of New World monkey) is due to a short-wavelength-sensitive (S) opsin gene on an autosome and a middle- (M) and a long-wavelength (L) opsin gene on the X chromosome¹. Other New World monkeys have only one autosomal and one X-linked opsin gene, but a polymorphism at the X-linked locus enables heterozygous female New World monkeys to be trichromatic^{1,5}, although males and homozygous females are dichromatic.

In contrast, nocturnal prosimians are thought to be monochromatic, as they have no functional autosomal opsin gene³.

Diurnal prosimians have a functional autosomal opsin gene and a functional X-linked opsin gene, but so far no polymorphism at either locus has been found². The spectral wavelength-sensitivity maxima ($\lambda_{\max}$) of opsins from four lemurs from each of two species have been measured by using electroretinographic flicker photometry²: only a single class of X-linked opsin was detected, with $\lambda_{\max}$ at about 543 nm, indicating that prosimians have no polymorphism at the X-linked opsin locus and are at best dichromatic². But because this conclusion was based on a small sample size, we studied 20 species representing the major prosimian lineages and did population screenings on several species by sequencing exons 3, 4 and 5 of the X-linked opsin gene.

These screenings revealed that there is an M/L polymorphism in both Coquerel's sifaka (*Propithecus verreauxi coquereli*) and the red ruffed lemur (*Varecia variegata rubra*) (Table 1), which represent the two major diurnal prosimian lineages. The M opsin is presumably of the same type as that found in diurnal lemurs and bushbabies^{2,3,6}, but the L opsin has not been detected before. The two alleles are quite common in both species, with the M allele having a frequency of 42% in the red ruffed lemur, for example, although this estimate has a large standard error. The M/L polymorphism is also found in a nocturnal prosimian, the greater dwarf lemur (*Cheirogaleus major*).

The X-linked opsin polymorphism and the autosomal opsin gene should enable a heterozygous female lemur to produce three classes of opsin cone, making it trichromatic in the same way as many New World monkeys. Behavioural and spectral studies of heterozygous female lemurs have yet to demonstrate trichromacy, but the possibility is supported by anatomical and physiological findings showing many similarities in the organization of prosimian and simian visual systems⁷, such as the parvocellular (P-cell) system, which is specialized for trichromacy by mediating red-green colour opponency⁸.

The phylogenetic distribution of the M and L opsins supports the idea that the X-linked opsin polymorphism and primate trichromacy arose early during primate evolution. Under the parsimony criterion, our observation that M and L opsin genes intermingle between species within each of the three major families of Madagascar lemurs (Indridae, Cheirogaleidae and Lemuridae) (Table 1), together with the finding of the polymorphism in some species, indicates that the ancestral strepsirrhine may have been polymorphic at the X-linked opsin locus and therefore trichromatic.

We found that the western tarsier (*Tarsius bancanus*) has an M opsin and that the Philippine tarsier (*T. syrichta*) has an L opsin, indicating that their diurnal common ancestor had an M/L polymorphism and was trichromatic. The common ancestor of tarsiers and strepsirrhines might therefore also have been trichromatic. Although the X-linked opsin polymorphism in prosimians may have arisen independently of that in New World monkeys, the P543 and P558 alleles of New World monkeys might have descended from the tarsiers.

Ying Tan, Wen-Hsiung Li

Department of Ecology and Evolution,
University of Chicago, 1101 East 57th Street,
Chicago, Illinois 60637, USA
e-mail: whli@uchicago.edu

Table 1 Distribution of the middle- and long-wavelength opsins among 20 species of prosimian

Family	Species	Samples		X chromosomes	Alleles	
		Male	Female		M	L
Loridae (N)	Slender loris	0	1	2	2	0
	Common slow loris	0	1	2	2	0
	Potto	1	0	1	1	0
Galagonidae (N)	Greater bushbaby	0	1	2	2	0
	Lesser bushbaby	1	1	3	3	0
	Aye-aye	0	3	6	6	0
Daubentonidae (N)	Coquerel's sifaka	9	10	29	4	25
	Golden-crowned sifaka	0	1	2	2	0
	Greater dwarf lemur	3	1	5	4	1
Indridae (D)	Coquerel's dwarf lemur	1	1	3	0	3
	Fat-tailed dwarf lemur	9	3	15	15	0
	Grey mouse lemur	0	2	4	0	4
Cheirogaleidae (N)	Red ruffed lemur	2	5	12	5	7
	Black/white ruffed lemur	4	1	6	6	0
	Brown lemur	1	0	1	1	0
Lemuridae (D)	Mongoose lemur	0	2	4	4	0
	Ring-tailed lemur	1	0	1	1	0
	Bamboo lemur	1	0	1	0	1
Tarsiidae (N)	Philippine tarsier	5	10	25	0	25
	Western tarsier	5	1	7	7	0

The identification of M and L opsins was based on amino acids at the residue sites responsible for spectral tuning. Previous studies^{9–12} have identified five critical sites in exons 3, 4 and 5 (amino-acid positions 180, 229, 233, 277 and 285) causing a spectral shift of ~5, 2, 1, 8 and 15 nm, respectively. The M opsin found in tarsiers and most Madagascan lemurs has the amino acids alanine, isoleucine, serine, tyrosine and alanine at these sites, which are identical to those of the marmoset and tamarin P543 allele^{12,13}. The M opsin found in bushbabies, lorises, potto and aye-ayes differs from the P543 allele only at position 229, having valine instead of isoleucine, consistent with a study of two bushbabies⁶. This difference is unlikely to affect spectral tuning^{9–12} as it does not involve a hydroxyl group, so $\lambda_{\max}$ is estimated for both M opsins to be 543 nm, the same as that of the marmoset P543 allele. The L opsin has alanine, isoleucine, serine, tyrosine and threonine at the five critical sites, which are identical to those of the P558 allele in the saki monkey, a New World monkey (Y. Tan and W.-H. Li, unpublished data), so its $\lambda_{\max}$ is estimated at about 558 nm. The three species polymorphic for the M and L alleles are shown in bold. N, nocturnal; D, diurnal.

- Jacobs, G. H. *Biol. Rev.* **68**, 413–471 (1993).
- Jacobs, G. H. & Deegan, J. F. II *Am. J. Primatol.* **30**, 243–256 (1993).
- Deegan, J. F. II & Jacobs, G. H. *Am. J. Primatol.* **40**, 55–66 (1996).
- Jacobs, G. H., Neitz, M., Deegan, J. F. II & Neitz, J. *Nature* **382**, 156–158 (1996).
- Travis, D. S., Bowmaker, J. K. & Mollon, J. D. *Vision Res.* **28**, 481–490 (1988).
- Zhou, Y.-H., Hewett-Emmett, D., Ward, J. P. & Li, W.-H. *J. Mol. Evol.* **45**, 610–618 (1997).
- Casagrande, V. A. & Kaas, J. H. in *Cerebral Cortex*, Vol. 10, *Primary Visual Cortex in Primates* (eds Peters, A. & Rockland, K. S.) 201–259 (Plenum, New York, 1994).
- Yamada, E. S., Marshak, D. W., Silverman, L. C. L. & Casagrande, V. A. *Vision Res.* **38**, 3345–3352 (1998).
- Neitz, M., Neitz, J. & Jacobs, G. H. *Science* **252**, 971–974 (1991).
- Merbs, S. L. & Nathans, J. *Photochem. Photobiol.* **58**, 706–710 (1993).
- Asenjo, A. B., Rim, J. & Oprian, D. D. *Neuron* **12**, 1131–1138 (1994).
- Shyue, S.-K. et al. *J. Mol. Evol.* **46**, 697–702 (1998).
- Hunt, D. M., Williams, A. J., Bowmaker, J. K. & Mollon, J. D. *Vision Res.* **33**, 147–154 (1993).

Nonlinear dynamics of lava dome extrusion

O. Melnik*† & R. S. J. Sparks*

* Centre for Environmental and Geophysical Flows, Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK

† Institute of Mechanics, Moscow State University, 1 Mitchurinski prosp, 117192, Moscow, Russia

During the eruption of the Soufrière Hills volcano, Montserrat (1995–99), and several other dome eruptions, shallow seismicity, short-lived explosive eruptions and ground deformation patterns indicating large overpressures (of several megapascals) in the uppermost few hundred metres of the volcanic conduit have been observed. These phenomena can be explained by the nonlinear effects of crystallization and gas loss by permeable flow, which are here incorporated into a numerical model of conduit flow and lava dome extrusion. Crystallization can introduce strong feedback mechanisms which greatly amplify the effect on extrusion rates of small changes of chamber pressure, conduit dimensions or magma viscosity. When timescales for magma ascent are comparable to timescales for crystallization, there can be multiple steady solutions for fixed conditions. Such nonlinear dynamics can cause large changes in dome extrusion rate and pulsatory patterns of dome growth.

Lava dome eruptions involve the ascent of gas-rich magmas, which lose gas during ascent¹. Degassing results in rheological stiffening^{2,3} due to gas exsolution and microlite crystallization from undercooled melt. Viscosity can increase by several orders of magnitude when silicic melts containing a few per cent dissolved water decompress and degas⁴. Microlite crystallization from undercooled melt accentuates the viscosity increase⁵, and can lead to the magma changing isentropically into a hot crystalline solid with mechanical strength. Crystallization of microlites and gas loss by permeable flow are time-dependent processes that introduce strong nonlinearities into the dynamics of conduit flow and dome extrusion. Here we develop a model of conduit and lava extrusion which incorporates these processes. The nonlinear behaviour identified in our calculations provide explanations for some important features observed in the eruption of the andesite lava dome of the Soufrière Hills, Montserrat, including ground deformation patterns, shallow seismicity, occurrence of short-lived vulcanian explosions, and large fluctuations of magma discharge rate. The model is, however, generic and can also explain phenomena at other eruptions, such as Mount Unzen in Japan.

Model development

Models of lava extrusion⁶ commonly involve application of the Poiseuille flow law. Here extrusion rate is inversely proportional to the viscosity, proportional to the conduit width raised to some power (4 for a cylinder), and proportional to the pressure gradient. But this model is too simple for the study of intermediate and silicic magmas, because of the rheological stiffening that accompanies degassing and the variable density of vesiculating magma. Many aspects of these complications have already been considered. Gas loss by permeable flow to the conduit walls can cause nonlinear variations in flow rate and transitions between explosive and extrusive activity^{7,8}. These transitions are very sensitive to small changes in parameter values, and the system of equations that incorporates gas loss can produce multiple solutions. Large viscosity variations related to gas loss cause strongly nonlinear pressure gradients and overpressures (defined here as the difference between magma pressure and local lithostatic pressure) at high levels in the conduit². Microlite crystallization can also potentially cause large overpressures to develop in the uppermost several hundred metres of conduits^{2,9}.

The physical framework for our model is shown in Fig. 1. Magma is stored in a chamber at depth L with a chamber overpressure, P_c , defined as chamber pressure minus lithostatic pressure, initial magma viscosity μ_0 and mass concentration of water dissolved in

the melt phase of c_0 . The magma ascends along a cylindrical conduit of diameter D and exits at the vent with pressure P_c . Flow in the dome is represented by a continuation of the conduit, with the same diameter for the active zone of flow within the dome and extrusion of new lava at the summit, consistent with observations at the Soufrière Hills volcano¹⁰. This simplification does not alter the underlying principles that emerge from our calculations. During ascent of the magma up the conduit and through the dome, gas is exsolved and is lost by vertical permeable flow through the magma column. In contrast to previous models^{7,8}, we consider only vertical gas flux. Although some gas loss to the conduit walls is inevitable, the observations of a strong gas plume emerging from the summit area of the Soufrière Hills dome¹¹ suggest that vertical flux is dominant. Sealing of the conduit walls also can prevent lateral gas loss^{12,13}.

We model the system shown in Fig. 1 by the following system of equations:

$$(1 - \alpha)(\rho_m(1 - \beta)(1 - c) + \rho_g\beta)V = Q_m \quad (1)$$

$$\rho_m(1 - \alpha)(1 - \beta)cV + \rho_g\alpha V_g = Q_g \quad (2)$$

$$V \frac{d\beta}{dz} = \beta^{2/3} \frac{\chi}{\mu_m} \left(1 - \frac{c}{c_0}\right); \quad \chi = 4\pi \left(\frac{3}{4\pi N}\right)^{2/3} \xi \quad (3)$$

$$\frac{dp}{dz} = -\rho g - \frac{32\mu V}{D^2} \quad (4)$$

$$V_g - V = -\frac{k(\alpha)}{\mu_g} \frac{dp}{dz} \quad (5)$$

$$\mu = \theta(\beta)\mu_m(c); \quad c = C_f \sqrt{p}; \quad p = \rho_g RT \quad (6)$$

$$\log[k(\alpha)/k_0] = -10.2(\alpha 10^2)^{1.4 \times 10^{-2} \alpha^{-1}}; \quad \alpha > 0.03 \quad (7)$$

$$\log[\theta(\beta)/\theta_0] = \arctan[\omega(\beta - \beta_*)] + \pi/2 \quad (8)$$

where ρ_m , ρ_c and ρ_g are the densities of melt, crystals and gas, respectively, ρ is the density of mixture, α and β are volume concentrations of respectively bubbles and crystals (in the condensed phase), V and V_g are velocities of magma and gas, respectively, Q_m and Q_g are discharge rates of respectively magma and gas per unit area, p is pressure, c is the mass concentration of dissolved gas, μ , μ_m and μ_g are viscosities of magma, melt and gas, respectively, D is the conduit diameter, C_f is the solubility constant,

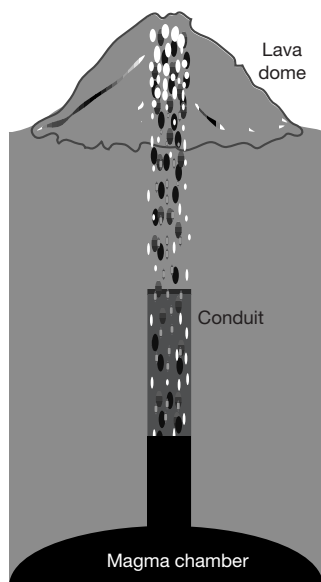


Figure 1 Schematic view of the volcanic system. Magma ascends from a chamber connected with the growing lava dome by a cylindrical conduit. Flow in the dome is represented by a continuation of the conduit, with extrusion of new lava at the summit of the dome. Gas is lost by permeable flow along the conduit near the surface.

$k(\alpha)$ is the permeability, R is the gas constant, T is temperature, z is the vertical coordinate, g is the acceleration due to gravity, N is the number density of crystals per unit volume, ξ is a constant related to crystal growth kinetics, and k_0 , ω , β_* , θ_0 are constants in permeability and viscosity laws as discussed below. Function $\theta(\beta)$ represents the influence of crystals on magma viscosity. Equations (1) and (2) represent conservation of mass for gas and condensed phase, and equation (3) describes crystal growth kinetics. Equation (4) states the conservation of momentum for the mixture in which the inertial term is negligibly small and conduit resistance is taken in Poiseuille form. Equation (5) is Darcy's law, and equations (6–8) summarize the physical properties of magma for isothermal flow conditions.

We recognize that this model is not a complete description, as there may still be important processes that have not been incorporated. The calculations, for example, assume that the ascent is sufficiently slow that the gas phase in the system is in equilibrium (that is, diffusion of gas into the bubbles is not rate-limiting). More elaborate models can be developed which incorporate disequilibrium in the gas phase, more realistic treatment of gas and magma movement in the dome and lateral pressure gradients in the conduit. To illustrate the principles of the model and to test the results on a well documented eruption, we have chosen parameters based on the Soufrière Hills andesite lava dome eruption (Table 1). We have also carried out sensitivity studies to establish how variations in important parameters listed in Table 1 affect the results.

We have made measurements of the matrix permeability and porosity of samples from the Soufrière Hills eruption with porosities in the range of 0 to 70%. These and other data¹⁴ can be described by equation (7) with most data falling between curves for $k_0 = 1$ and $k_0 = 10$. Most of the porosity consists of interconnected vesicles in the residual glass phase, which is continuously distributed between crystalline phases. We consider, in agreement with the conclusions of Klug and Cashman¹⁴, that there have been only minor changes in permeability due to cooling. As lava domes are highly fractured, the actual permeability in the upper parts of the conduit and dome may be even higher so we consider $k_0 = 100$ in addition to the impermeable end-member case ($k_0 = 0$). We will show *a posteriori* that the influence of permeable gas loss on extrusion rate is small for a wide range of k_0 .

Table 1 Parameters for the Soufrière Hills andesite eruption used in modelling

Parameter	Symbol	Value range	Information sources
Magma chamber depth	L	5 km	Earthquakes ³² and phase equilibria ³³
Magma chamber pressure	P_c	0–20 MPa	Range of crustal strengths
Melt water content	c_0	5%	Ref. 33
Magma temperature	T	850 °C	Refs 25, 33
Magma crystal content	β_0	0.6	Ref. 25
Conduit diameter	D	30 m	Dimensions of spines and early crater; hornblende reaction rims ³⁴
Viscosity coefficient	θ_0	1.6	From calculations
Viscosity coefficient	β_*	0.62	From calculations
Viscosity coefficient	ω	20.6	From calculations
Crystal growth rate coefficient	χ	$7 \times 10^{-3} \text{ s}^{-1}$	From calculations
Permeability coefficient	k_0	1	Regression through data
Density of melt	ρ_m	$2,300 \text{ kg m}^{-3}$	
Density of crystals	ρ_c	$2,700 \text{ kg m}^{-3}$	
Density of wallrock	ρ_r	$2,600 \text{ kg m}^{-3}$	
Solubility coefficient	C_f	$4.1 \times 10^{-6} \text{ Pa}^{1/2}$	
Gas viscosity	μ_g	$1.5 \times 10^{-5} \text{ Pa s}$	

The influence of crystals on viscosity is given by equation (8), and an empirical formula which reproduces the expected shape of the relationship between crystal content and viscosity across the transition region where crystals become close-packed, as observed in experiments⁵.

We have chosen a simple parameterization of crystal growth (equation (3)) based on theoretical and experimental studies^{15,16} in which growth rate is proportional to undercooling and inversely proportional to melt viscosity. Both degree of undercooling and melt viscosity are dependent on dissolved gas content, and are thus strongly dependent on pressure. We assume for simplicity that there is a single crystal nucleation event, that is, N and χ are constants in equation (3). For magma viscosity we have taken the relevant magma parameters for the Soufrière Hills andesite (Table 1), and have applied experimentally based algorithms^{4,5}. Very large changes in magma viscosity occur when the crystal content reaches a threshold value at the condition of close packing⁵. There are considerable uncertainties in assessing the appropriate threshold crystal content for natural magmas, as the melt fraction at the close-packed condition is dependent on both crystal shape⁵ and grain size distribution¹⁷.

There are two remaining problems in choosing parameters: that the crystal growth parameter, χ , and the viscosity coefficient $\theta(\beta)$, are not well constrained. Although there are some experimental data on crystal growth, none of the available data are sufficiently close to a natural system to provide a robust value of χ . The high crystal content of the Soufrière Hills magma in the chamber (estimated at 60–65%) means that the magma is close to conditions of close-packing. There are no experimental data that can help constrain this parameter accurately, yet the results of calculations are likely to be sensitive to the choice of $\theta(\beta)$ in equation (8). We therefore use observations of dome height, extrusion rate, and crystal content from the Soufrière Hills to calibrate values of χ and $\theta(\beta)$ to provide the basis for a more extensive investigation of the effects of the nonlinearities on flow. We note that the choice of rheological parameters are such that the viscosity of the Soufrière Hills andesite has a variation consistent with independent estimates made from petrological models⁴ and observations of dome growth rates.

Feedback mechanisms are evident in the mathematical description. Lower magma permeability results in a decrease in the overall

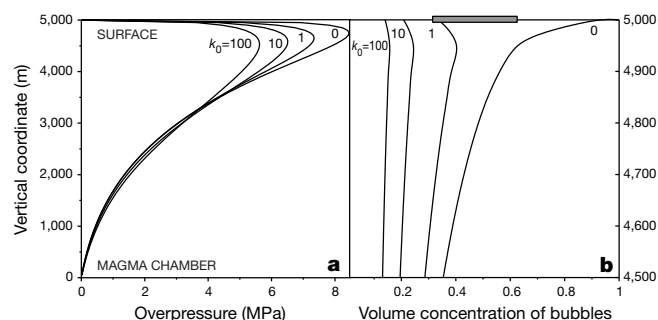


Figure 2 Calculated profiles of overpressure (a) and porosity (b) along a conduit. Conditions for calculations are described in the text. Overpressure (the difference between the magma pressure and lithostatic pressure) reaches a maximum in the uppermost few hundred metres of a volcanic conduit. Curves are shown for values of the permeability coefficient k_0 from values of 0 to 100. The range of observed porosities in pumice clasts from vulcanian eruptions of the Soufrière Hills is shown as a horizontal bar at the top of **b**. These data suggest that values of k_0 from 1 to 0.1 are typical. Low values of k_0 give unrealistically high porosities at the top of the conduit and explosive conditions are likely to develop in practice. Porosity profiles are shown for the upper 500 m of the conduit. The calculations here assume no crystallization in the conduit ($\chi = 0$).

weight of the magma due to higher amounts of retained gas. The excess pressure driving magma from the chamber, defined as the difference between the chamber pressure and magma column weight, must consequently increase. As magma permeability increases, a larger proportion of the gas escapes, magma column weight increases, and the excess driving pressure decreases. Another example is the case when the timescale for magma extrusion is comparable to the timescale for microlite crystallization. If there is a slight change in some parameter that enhances flow rate (such as chamber pressure, conduit width or initial magma viscosity), then gas loss and crystallization will amplify the effect. An increase in flow rate will give less time for crystal growth, and so reduce overall viscosity. Also, an increase in flow rate gives less time for gas loss, therefore the magma column weight decreases. According to equation (4), the magma ascent velocity will consequently increase. We will show below that these feedback amplifications can be substantial.

Results of calculations and comparison with observations

We first demonstrate the influence of the permeability coefficient on eruption dynamics. Figure 2 illustrates the role of gas loss and permeability on flow dynamics, in terms of porosity and overpressure variations with depth. The calculations show a region of high porosity and high overpressure with a maximum in the uppermost parts of the conduit, at typical depths of a few hundred metres below the vent. The maximum overpressure depends on the permeability values, with a range of 4 to 8 MPa.

The model can explain, in the case of the Soufrière Hills, ground deformation patterns that indicate pressurization of the conduit rather than the deep magma chamber. Shepherd *et al.*¹⁸ observed that deformation during 1996 could be described as a simple decrease in the velocity of ground movements radially away from the dome. They inferred a pressure source at 700 m depth with an estimated overpressure of 10 MPa. Voight *et al.*³ estimated a pressure source at 400 m depth from cyclic patterns of ground inflation and deflation. Dome eruptions are characterized by shallow seismicity with long-period components^{2,19,20}, and by short-lived (a few tens of seconds) vulcanian explosions from shallow pressurized sources^{3,9,20,21}. These phenomena can be explained by the predicted gas overpressures exceeding magma and conduit wallrock strength. Explosions occur when internal gas pressures exceed magma strength²². During the Soufrière Hill eruption, there were 85 vulcanian explosions in the August to October 1997 period²³. Volume constraints indicate that these explosions originated from

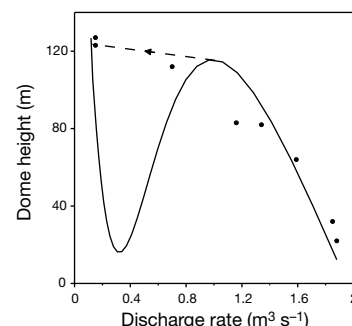


Figure 3 Calculated and observed magma discharge rate versus dome height. Shown are data for the October 1997 lava dome growth at Soufrière Hills volcano (filled circles), and the results of numerical calculations using equations (1)–(8). The dashed line shows the expected unsteady path for the system to change from one steady state to another with declining flow rate at approximately constant dome height: here the system cannot follow the steady path, because this would require a large decrease in dome height which is unphysical.

the uppermost few hundred metres of the conduit. Samples of the pumice had angular platy shapes indicative of brittle fragmentation and porosities of 30–60%, indicating a zone of high porosity and overpressures of a few megapascals, as predicted in the model. The calculated overpressures of several megapascals are comparable to the strength of the crystalline magma and wallrocks²², so that the system is inherently unstable to perturbations such as unloading of the pressurized conduit by dome collapse. The cyclic patterns of seismicity, ground deformation, and eruptive activity documented by Voight *et al.*³ can be seen as the consequence of pressurization repeatedly building up the system to failure conditions.

We now explore the effects of groundmass crystallization. Pumice and samples of the Soufrière Hills dome that were erupted during periods of high discharge²⁴ have high glass contents (25–35%) and few microlites²⁵, whereas samples derived from parts of the dome that were extruded more slowly (typically, days to weeks) have much lower glass contents (5–15%) and high contents of groundmass microlites. These, and other, observations^{26–28} suggest that microlite crystallization can take place on similar timescales to the ascent time of the magma (see Table 1).

Figure 3 shows (as data points) the observed magma discharge rate versus dome height at the Soufrière Hills during October 1996. We first attempted to model the observations with flow models that did not include crystallization kinetics. Using data in Table 1 we were able to fit the initial discharge rate of $1.8 \text{ m}^3 \text{ s}^{-1}$, but found that the ultimate height of the dome was much greater than observed (our best-fit value being about 600 m compared to 120 m). The height/discharge rate relationship could not be reconciled with the initial discharge rate at zero dome height: we therefore manipulated the controlling variables, and found values of χ and the constants in $\theta(\beta)$ that were consistent with the observations. The resulting calculated curve is shown in Fig. 3, and the parameter values are listed in Table 1. The model predicts strongly nonlinear behaviour when the height of the dome growth reaches about 120 m. At this point, dome growth cannot then follow the steady solution path as this requires a substantial decrease in height. Rather, we expect the dome to follow an unsteady path at approximately constant dome height until the curve for the steady solution is once more met at very low flow rate (Fig. 3). Large changes in extrusion rate have been observed with minor changes in dome height both at Soufrière Hills²⁴ and at Mount Unzen²⁰.

Figure 4 shows the relationship between extrusion rate and magma chamber overpressure, using values of χ and $\theta(\beta)$ determined from the dome growth in October 1997 (Fig. 3) and

contoured in terms of dome height from 0 to 300 m. The extrusion rate and heights cover the same range as observed at the Soufrière Hills eruption²⁴. Figure 4b shows only the curves for 0 and 300 m dome height, and illustrates possible paths for eruption behaviour. Above a magma chamber pressure of 10 MPa the variations are monotonic, and the numerical results are close to the simple Poiseuille flow law. Extrusion rate decreases with dome height. In this region flow rates are too fast to allow significant crystallization. The slight departure from simple Poiseuille flow is due to the incorporation of gas loss by permeable flow.

For magma chamber overpressures below about 10 MPa the curves become more complex and sigmoidal in shape, indicating the possibility of multiple steady solutions for fixed eruption conditions. In this region, microlite crystallization occurs during ascent and introduces the strong feedback mechanisms described above. For low chamber overpressures (P_c) small changes in P_c can induce large changes in flow rate. We found similar strong sensitivity to small changes in conduit diameter or initial magma viscosity. For example, a change in viscosity or chamber pressure by a factor of two, or a 20% change of conduit width, can cause over an order of magnitude change in extrusion rate. The Poiseuille flow law would only result in a factor of 2 change.

The multiple steady solutions allow for complex pulsatory behaviour (Fig. 4b). A dome starting to grow at point A may reach 300 m height at point B. Beyond that height, the steady solution drops down to a much lower extrusion rate at point C. If the chamber pressure increases (for example due to magma replenishment⁶), then the path C, D, E can be followed and this might lead to periodic behaviour as the dome jumps between one steady state and another. A significant dome collapse event can decrease dome height suddenly, triggering greatly accelerated dome growth rates as shown in path E to F (Fig. 4).

Observations of dome growth at the Soufrière Hills^{3,24} and Mount Unzen²⁰ showed pulsations on a wide range of timescales, from hours to months, with extrusion rates varying by a factor of more than 50. The composition of the magmas did not vary during the course of these eruptions, so variations in the initial magma crystal content and viscosity in the chamber are unlikely to have been a major cause of these pulsations and the large variations in extrusion rate. Relatively minor variations in the conduit dimensions and magma chamber pressure could have been amplified by the nonlinear effects to cause the wide range of extrusion rate and marked fluctuations.

The model we present here is limited by the need to calibrate it to a particular dome eruption; this is required to estimate poorly

constrained parameters relating to the rheological properties of high-crystal-content magmas and crystallization kinetics. Although an independent estimate of the values of these parameters would be preferable, the choice of calibrated values allows us to learn about the general principles of nonlinear dynamics using values which are at least consistent with observations.

Implications for monitoring and forecasting

Our model of conduit flow and dome extrusion incorporates the nonlinear effects of degassing and crystallization. The calculations provide a mechanism for generating large overpressures at shallow levels in such eruptions. Evidence for overpressured conditions and pulsatory activity as characteristic of dome eruptions comes not only from the Soufrière Hills, but from elsewhere, such as Mount Unzen in Japan²⁰, Santiaguito in Guatemala²⁹, Lascar in Chile²¹, Galeras in Columbia⁹, and Mount St Helens in the USA³⁰. The flow model introduces effects that help to explain ground deformation, shallow seismicity and short-lived vulcanian explosions. Cyclic patterns of behaviour, as at the Soufrière Hills³, can be explained by build-up of overpressure in the upper parts of a conduit to critical threshold conditions for magma flow or explosive activity. Other mechanisms, such as stress weakening with stick-slip behaviour³¹, may also play a role in short timescale cycles.

The model has important implications for the forecasting of volcanic eruptions. It suggests that many of the geophysical signals monitored in dome eruptions relate to pressurization in the upper parts of the conduit rather than the deep magma chamber. This link between flow models of degassing magma and monitoring data provides a new framework for interpretation of geophysical data.

Nonlinear dynamical effects allow multiple solutions to exist for fixed eruption parameters. Additionally, these eruptive systems can be extremely sensitive to minor changes in magma properties, magma chamber conditions and conduit dimensions. Unsteady transitions between stable states are themselves likely to prove of considerable interest. Nonlinear and unsteady behaviour makes possible transitions to chaotic behaviour as the system attempts to achieve equilibrium. Under some circumstances behaviour can become inherently unpredictable. Improved knowledge of crystal growth kinetics, gas exsolution kinetics (not incorporated here), permeability development in vesiculating magmas and rheological properties of crystal-rich magmas will help develop more complex and realistic models. However, nonlinearity will inevitably feature in future models, no matter how well constrained the eruption parameters and physical magma properties are. □

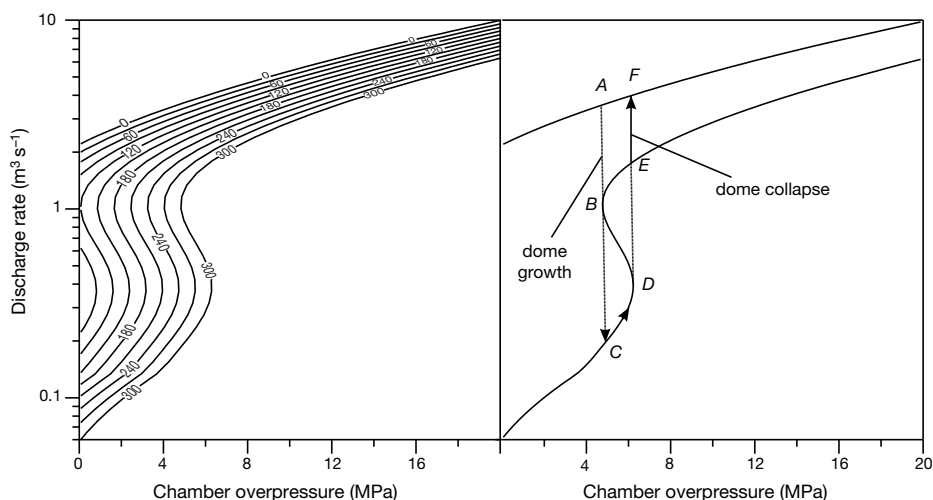


Figure 4 The relationship between extrusion rate and magma chamber pressure (a) and possible paths of unsteady eruption (b). Contours in a are of dome height from 0 to 300 m. See text for details about points A–F.

Received 4 June; accepted 27 September 1999.

1. Eichelberger, J. C., Carrigan, C. R., Westrich, H. R. & Price, R. H. Non-explosive silicic volcanism. *Nature* **323**, 598–602 (1986).
2. Sparks, R. S. J. Causes and consequences of pressurisation in lava dome eruptions. *Earth Planet. Sci. Lett.* **150**, 177–189 (1997).
3. Voight, B. *et al.* Magma flow instability and cyclic activity at Soufriere Hills Volcano, Montserrat, B.W.I. *Science* **283**, 1138–1142 (1999).
4. Hess, K. U. & Dingwell, D. B. Viscosities of hydrous leucogranite melts: a non-Arrhenian model. *Am. Mineral.* **81**, 1297–1300 (1996).
5. Lejeune, A. M. & Richet, P. Rheology of crystal-bearing melts: an experimental study at high viscosities. *J. Geophys. Res.* **100**, 4215–4229 (1995).
6. Stasiuk, M. V., Jaupart, C. & Sparks, R. S. J. Variations of flow rate and volume during eruption of lava. *Earth Planet. Sci. Lett.* **114**, 505–516 (1993).
7. Jaupart, C. & Allegre, C. J. Gas content, eruption rate and instabilities of eruption regime in silicic volcanoes. *Earth Planet. Sci. Lett.* **102**, 413–429 (1991).
8. Woods, A. W. & Koyaguchi, T. Transitions between explosive and effusive eruption of silicic magmas. *Nature* **370**, 641–645 (1994).
9. Stix, J. *et al.* A model for vulcanian eruptions at Galeras volcano, Colombia. *J. Volcanol. Geotherm. Res.* **77**, 285–304 (1997).
10. Young, S. R. *et al.* Overview of the Soufriere Hills Volcano and the eruption. *Geophys. Res. Lett.* **25**, 3389 (1998).
11. Young, S. R. *et al.* Monitoring SO₂ emission at the Soufriere Hills volcano: implications for changes in eruptive conditions. *Geophys. Res. Lett.* **25**, 3681–3684 (1998).
12. Boudon, G., Villemant, B., Komorowski, J. C., Idefonse, P. & Semet, M. P. The hydrothermal system at the Soufriere Hills Volcano, Montserrat (West Indies): Characterisation and role in the ongoing eruption. *Geophys. Res. Lett.* **25**, 3693–3696 (1998).
13. Fink, J. H., Anderson, S. W. & Manley, C. R. Textural constraints on effusive silicic volcanism. *J. Geophys. Res.* **97**, 9073–9083 (1992).
14. Klug, C. & Cashman, K. V. Permeability development in vesiculating magmas: implications for fragmentation. *Bull. Volcanol.* **58**, 87–100 (1996).
15. Dowty, E. in *Physics of Magmatic Processes* (ed. Hargreaves, R. B.) 419–485 (Princeton Univ. Press, 1980).
16. Lofgren, G. in *Physics of Magmatic Processes* (ed. Hargreaves, R. B.) 485–551 (Princeton Univ. Press, 1980).
17. Sohn, H. Y. & Moreland, C. The effect of particle size distribution on packing density. *Can. J. Chem. Eng.* **46**, 162–167 (1968).
18. Shepherd, J., Herd, R. A., Jackson, P. & Watts, R. Ground deformation measurements at the Soufriere Hills Volcano, Montserrat, 2, Rapid static GPS measurements. June 1996–June 1997. *Geophys. Res. Lett.* **25**, 3413–3416 (1998).
19. Miller, A. D. *et al.* Seismicity associated with dome growth and collapse of the Soufriere Hills Volcano, Montserrat. *Geophys. Res. Lett.* **25**, 3401–3405 (1998).
20. Nakada, S., Shimizu, H. & Ohta, K. Overview of the 1990–1995 eruption at Unzen Volcano. *J. Volcanol. Geotherm. Res.* **89**, 1–22 (1999).
21. Matthews, S. J., Gardeweg, M. C. & Sparks, R. S. J. The 1984–1996 cyclic activity of Lascar volcano, Northern Chile: cycles of dome growth, dome subsidence, degassing and explosive eruptions. *Bull. Volcanol.* **59**, 72–82 (1997).
22. Adilbirov, M. & Dingwell, D. B. High temperature fragmentation of magma by rapid decompression. *Nature* **380**, 146–149 (1996).
23. Young, S. R. *et al.* *Dome Collapse and Vulcanian Explosive Activity, September to October 1997* 1–23 (Special Rep. No. 5, Montserrat Observatory, Mingo Hill, Montserrat, 1999).
24. Sparks, R. S. J. *et al.* Magma production and growth of the lava dome of the Soufriere Hills volcano, Montserrat: November 1995 to December 1997. *Geophys. Res. Lett.* **25**, 3421–3424 (1998).
25. Murphy, M. D., Sparks, R. S. J., Barclay, J., Carroll, M. R. & Brewer, T. S. Remobilisation origin for andesite magma by intrusion of mafic magma at the Soufriere Hills Volcano, Montserrat, W.I., a trigger for renewed eruption. *J. Petrol.* (in the press).
26. Cashman, K. V. Groundmass crystallization of Mount St Helens dacite, 1980–1986: a tool for interpreting shallow magmatic processes. *Contrib. Mineral. Petrol.* **109**, 431–449 (1992).
27. Nakada, S. & Motomura, Y. Petrology of the 1991–1995 eruption at Unzen: effusion pulsation and groundmass crystallization. *J. Volcanol. Geotherm. Res.* **89**, 173–196 (1999).
28. Hammer, J. E., Cashman, K. V., Hoblitt, R. P. & Newman, S. Degassing and microlite crystallization during pre-climatic events of the 1991 eruption of Mt. Pinatubo, Philippines. *Bull. Volcanol.* **60**, 355–380 (1999).
29. Rose, W. I. Pattern and mechanism of volcanic activity at Santiaguito volcanic dome, Guatemala. *Bull. Volcanol.* **37**, 73–94 (1973).
30. Swanson, D. A. & Holcomb, R. T. *Lava Flows and Domes: Emplacement Mechanisms and Hazards Implications* (ed. Fink, J. H.) 3–24 (Springer, Berlin, 1990).
31. Denlinger, R. & Hoblitt, R. P. Cyclic eruptive behaviour of silicic volcanoes. *Geology* **27**, 459–462 (1999).
32. Aspinall, W. P. *et al.* Soufriere Hills eruption, Montserrat: 1995–1997: volcanic earthquake locations and fault plane solutions. *Geophys. Res. Lett.* **25**, 3397–3400 (1998).
33. Barclay, J. *et al.* Experimental phase equilibria: constraints on pre-eruptive storage conditions of the Soufriere Hills magma. *Geophys. Res. Lett.* **25**, 3437–3440 (1998).
34. Devine, J. D., Rutherford, M. J. & Gardner, J. C. Petrologic determination of ascent rates for the 1995–1997 Soufriere Hills Volcano andesite magma. *Geophys. Res. Lett.* **25**, 3673–3676 (1998).

Acknowledgements

We thank H. Huppert, H. Mader and A. Woods for comments, and J. Fink and K. Cashman for criticisms and suggestions. This work was supported by the NERC, the Russian Foundation for Basic Research, and an NERC fellowship (R.S.J.S.).

Correspondence and requests for materials should be addressed to R.S.J.S. (e-mail: Steve.Sparks@bris.ac.uk).

Lower Cambrian vertebrates from south China

D-G. Shu*, H-L. Luo†, S. Conway Morris‡, X-L. Zhang*, S-X. Hu†, L. Chen*, J. Han*, M. Zhu§, Y. Li* & L-Z. Chen†

* Early Life Institute and Department of Geology, Northwest University, Xi'an, 710069, People's Republic of China

† Yunnan Institute of Geological Sciences, 131 Baita Road, Kunming, People's Republic of China

‡ Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK

§ Institute of Vertebrate Palaeontology & Palaeoanthropology, Chinese Academy of Sciences, Beijing, 100044, People's Republic of China

The first fossil chordates are found in deposits from the Cambrian period (545–490 million years ago), but their earliest record is exceptionally sporadic and is often controversial. Accordingly, it has been difficult to construct a coherent phylogenetic synthesis for the basal chordates. Until now, the available soft-bodied remains have consisted almost entirely of cephalochordate-like animals from Burgess Shale-type faunas. Definite examples of agnathan fish do not occur until the Lower Ordovician (~475 Myr BP), with a more questionable record extending into the Cambrian. The discovery of two distinct types of agnathan from the Lower Cambrian Chengjiang fossil-Lagerstätte is, therefore, a very significant extension of their range. One form is lamprey-like, whereas the other is closer to the more primitive hagfish. These finds imply that the first agnathans may have evolved in the earliest Cambrian, with the chordates arising from more primitive deuterostomes in Ediacaran times (latest Neoproterozoic, ~555 Myr BP), if not earlier.

The 'Cambrian explosion' refers to the seemingly abrupt appearance of diverse metazoan groups, representing a number of extant phyla as well as some more problematic clades^{1,2}, during the Cambrian period. Faunas of the type found at the Burgess Shale¹ provide exceptional information on Cambrian marine communities. Recent documentation of their faunal riches has led to a series of provocative phylogenetic analyses^{3,4} which may be broadly congruent with the division of the protostomes into the Ecdysozoa and Lophotrochozoa⁵. However, the early evolution of the third triploblast superclade, the deuterostomes, is much less well understood.

In particular, the fossil record of Cambrian chordates is extremely meagre. The soft-bodied remains^{6–8} consist almost exclusively of cephalochordate-like animals, although they seem to be signifi-

cantly different from the Recent amphioxus. A unique and as yet unnamed specimen from the Burgess Shale has been compared to an ammocoete lamprey^{9,10}. Records of fish are confined to questionable scales¹¹ from the Upper Cambrian. The euconodonts, which are also chordates¹², also appear in the uppermost Cambrian, but their precise placement relative to other agnathans remains controversial^{12,13}. Transitional forms between euconodonts and paraconodonts¹⁴ imply that this group of chordates evolved during the Middle Cambrian, but nothing is yet known of the soft-part anatomy of the early representatives of this group.

Here we describe two distinct types of agnathan from the Lower Cambrian Chengjiang fossil-Lagerstätte of the Qiongzhusi Formation at Haikou¹⁵, Kunming City, Yunnan (Fig. 1). The more

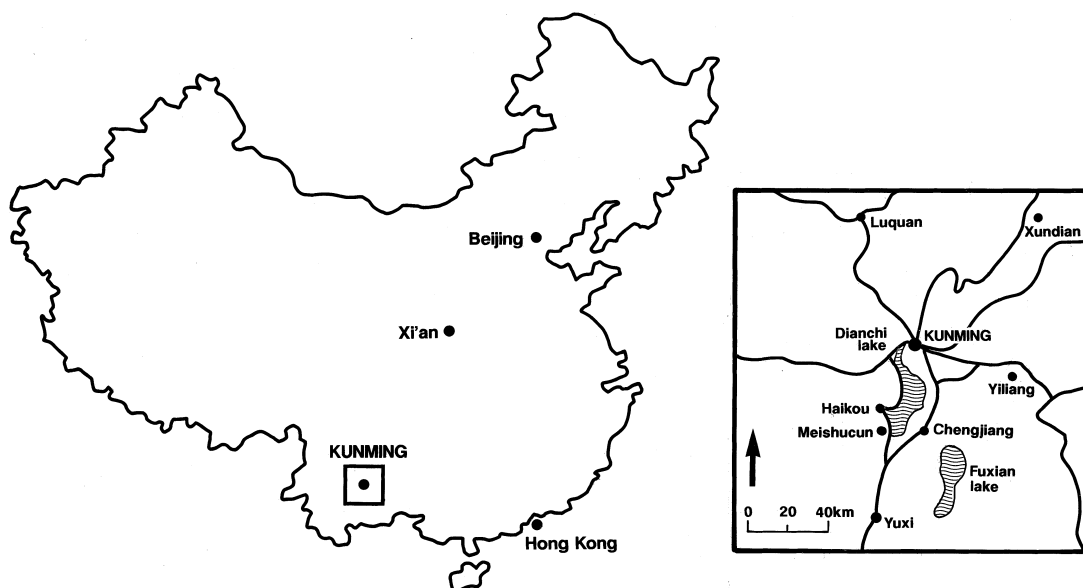


Figure 1 Map of Chengjiang area, Yunnan province and its position relative to the rest of China. The first discoveries of soft-bodied fossils were made in about 1910 near Yiliang¹, but the principal focus of activity has been near Chengjiang, especially the localities a few

kilometres north-east of Fuxian Lake, most notably Maotianshan. The vertebrate fossils described here come from close to Haikou, near Dianchi Lake.

primitive example, *Mylokunmingia* gen. nov., has well developed gill pouches with probable hemibranchs, whereas *Haikouichthys* gen. nov. has structures that we interpret as part of a branchial basket and a dorsal fin with prominent fin-radials. Shared features include complex myomeres, as well as probable paired ventral fin-folds and a pericardic cavity. These discoveries significantly predate previously published reports, but they also imply that even more primitive vertebrates had evolved before the mid-Lower Cambrian.

Phylum Chordata
Subphylum Vertebrata
Class Agnatha

Mylokunmingia Shu, Zhang & Han gen. nov.

Mylokunmingia fengjiao Shu, Zhang & Han sp. nov.

Etymology. The generic name refers to myllos (Greek, fish) and the capital of Yunnan. Species name from Chinese fengjiao, beautiful.

Holotype. Early Life Institute, Northwest University, Xi'an: ELI-0000201.

Stratigraphy and locality. Qiongzhusi (Chiungchussu) Formation, Yu'an-shan Member (*Eoredlichia* Zone); Lower Cambrian. Specimen collected from Haikou, Kunming City, Yunnan.

Diagnosis. Fusiform, divided into head region and trunk. Dorsal fin towards anterior, ventro-lateral fin-fold arising from trunk, probably paired. No fin-radials. Head bears series of five, possibly six, gill pouches, each with hemibranchs. Gill pouches possibly linked to extrabranchial atria. Trunk with ~25 myomeres, double V structure with ventral V directed posteriorly and dorsal V anteriorly. Internal anatomy includes alimentary canal with pharynx and intestine, notochord, and possible pericardic cavity.

Description. *Mylokunmingia* gen. nov. is fusiform, with a total length of about 28 mm and maximum height (excluding the dorsal fin) of 6 mm; this point is located about 11 mm from the anterior (Figs 2a, 3). Dermal skeleton or scales appear to have been absent. The body has two regions: a head, taken here to include both snout and complex structures interpreted as gills, and a segmented trunk.

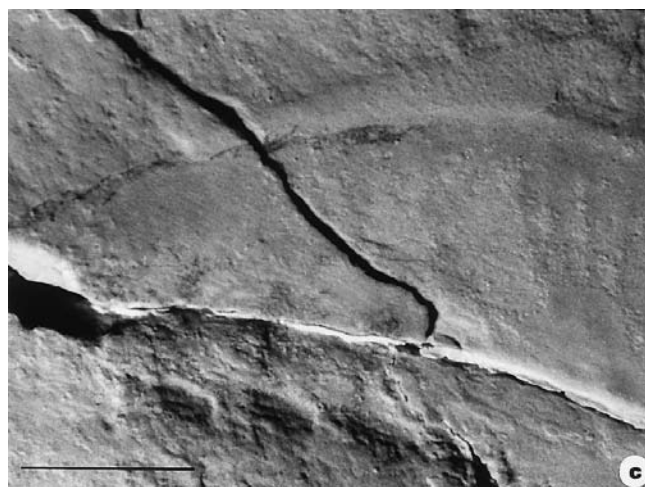
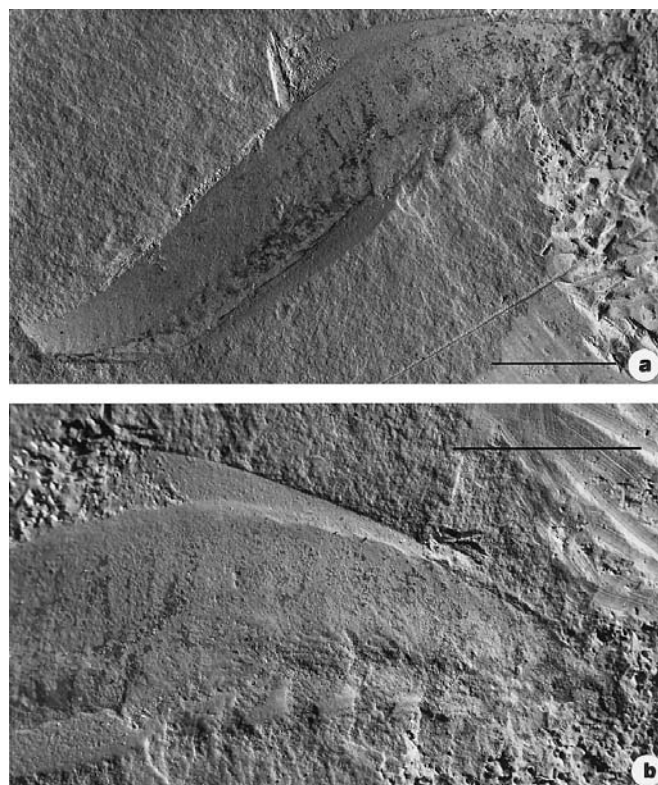


Figure 2 The Lower Cambrian agnathan vertebrate *Mylokunmingia fengjiao* Shu, Zhang & Han gen. et sp. nov. from Haikou, Yunnan. Specimen ELI-0000201. **a**, Entire specimen, anterior end to the right, posterior tip incomplete owing to post-mortem folding; scale bar equivalent to 5 mm. **b**, Detail of the head region of the fossil, to emphasize structures interpreted as gill pouches; scale bar equivalent to 5 mm. **c**, Anterior of counterpart, to show series of ventral structures interpreted as possible extrabranchial atria; scale bar equivalent to 3 mm.

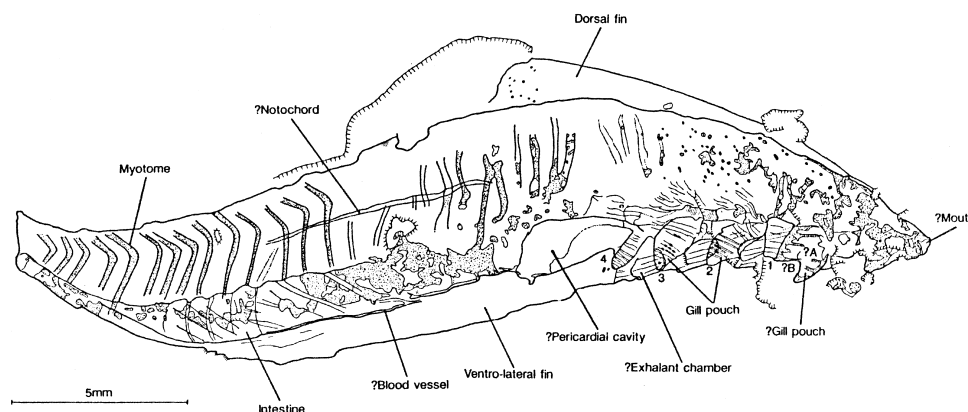


Figure 3 Camera-lucida drawing of specimen, with certain features (notably structures interpreted as extrabranchial atria) combined by reversal from the counterpart (Fig. 1c), to show interpretation. Numbers 1–4 refer to gill pouches identified with reasonable

certainty; ?A and ?B refer to more tentative identification of gill pouches, of which ?A is less certain than ?B.

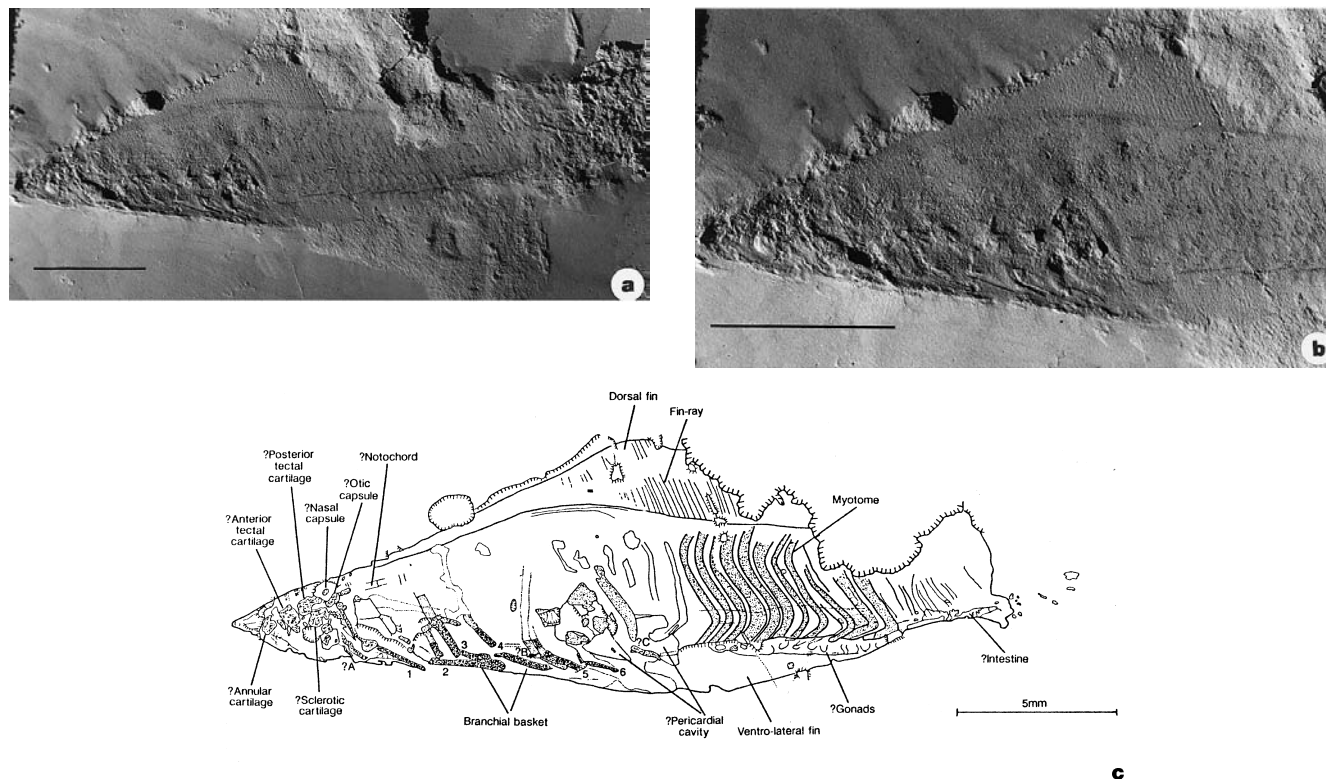


Figure 4 The Lower Cambrian agnathan vertebrate *Haikouichthys ercaicunensis* Luo, Hu & Shu gen. et sp. nov. from Haikou, Yunnan. Specimen HZ-f-12-127. **a**, Entire specimen, anterior to the left; more posterior region appears to fade out into sediment, possibly representing decay of body; attempts to excavate this area were not successful. Scale bar equivalent to 5 mm. **b**, Detail of anterior to show putative gill bars, possible elements of cranial endoskeleton, and pericardic area; scale bar equivalent to 5 mm. **c**, Camera-

lucida drawing of specimen to show interpretation. Numbers 1–6 indicate units of the branchial basket that are identified with some confidence; ?A–?C refer to less secure identifications. Two possible areas representing the pericardic cavity are indicated. To the anterior of ?C a triangular area with patches of diagenetic mineralization is one possibility; a fainter region to the posterior is the alternative location.

A sail-like dorsal fin arises fairly close to the anterior tip and increases to a maximum observed height of 1.5 mm. Its more posterior development is uncertain. On the ventro-lateral side, and more to the posterior, there is a prominent fin-fold which in the fossil is inclined at an angle to the rest of the trunk. This deflection may indicate that originally the fin-fold was paired, with the corresponding structure concealed at a deeper level within the sediment. Neither the dorsal nor the exposed ventral fin-fold display fin-radials.

The details of the head are complex (Fig. 2b, c). Most conspicuous is a series of relatively large, posteriorly inclined structures with transverse lineations. We interpret these as gill pouches with hemibranchs. Some of the latter have a beaded appearance, possibly indicating original folding or crenulation. Four pouches are clearly visible, the posteriormost being reduced in size. Anteriorly a fifth pouch is identified with some confidence, and more tentatively a sixth. Ventral to these pouches, and more clearly preserved on the counterpart (Fig. 2c), are convex (in the part) units, each with two or three longitudinal ridges. The position of these structures relative to the gill pouches is consistent with their representing extrabranchial atria (Fig. 3). The posteriormost pouch, however, appears to have lacked a well developed atrium. No branchial basket or other skeletal units have been identified in association with the gills. Other regions of the head show scattered patches of darker tissue. Apart from some segmental structures, close to the boundary with the trunk, these patches show no coherent pattern such as might represent parts of an endoskeleton. The mouth was presumably located at the anterior end and, although there is perhaps the suggestion of an annular ring (Fig. 2a, c), precise details are difficult to discern. An approximately oval area posterior to the gills may represent the pericardic cavity.

The trunk bears prominent sigmoidal segments (Figs 2a, 3), interpreted as myocommata, which in life separated about 25 myomeres. In lateral view the dorsal V points anteriorly and the ventral one posteriorly. Accordingly, on the dorsum the myomeres of either side would have converged posteriorly, and along the venter in an anterior direction. Towards the ventral margin of the trunk a dark area with positive relief presumably represents a sediment-filled intestine. The gut can be traced, more clearly in the counterpart, to the incomplete posterior termination. It is uncertain, therefore, whether *Mylokunmingia* possessed a post-anal tail. In the mid-region of the trunk there is a fairly prominent strand (Figs 2a, 3), which probably represents the incomplete remains of the notochord. At the trunk–fin-fold boundary there are very narrow dark lines, which show at least one bifurcation. These may be blood vessels.

Haikouichthys Luo, Hu & Shu gen. nov.

Haikouichthys ercaicunensis Luo, Hu & Shu sp. nov.

Etymology. After a locality at Haikou, near Kunming City, and the Greek ichthys. The specific name refers to Ercaicun, a small village close to where the specimen was found.

Holotype. Yunnan Institute of Geological Sciences, Kunming: HZ-f-12-127.

Stratigraphy and locality. As for *Mylokunmingia fengjiao*.

Diagnosis. Fusiform, more slender than *Mylokunmingia*, divided into head region and trunk. Prominent dorsal fin towards anterior, with fin-radials. Ventro-lateral fin-fold arising from trunk, probably paired. Head bears at least six, possibly up to nine, gill arches. Trunk with myomeres, double V as in *Mylokunmingia*. Internal anatomy includes cranial cartilages, pericardic cavity, intestine and metameric ?gonads, the last being arranged along the ventral trunk.

Description. *Haikouichthys* has a known length of about 25 mm

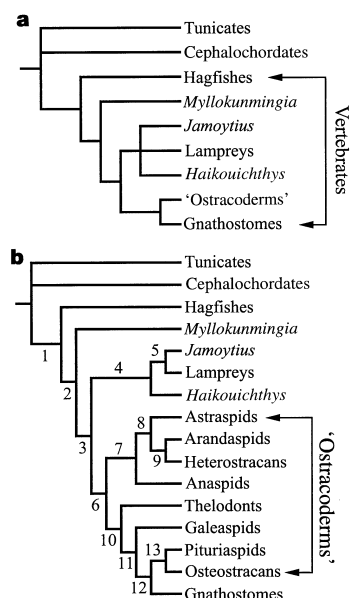


Figure 5 Phylogenetic analysis of early vertebrates, to include *Mylokunmingia* gen. nov. and *Haikouichthys* gen. nov. **a**, Strict consensus tree of the six most parsimonious trees. **b**, One of the six most parsimonious trees; the other five should show minimal differences. Data set based on refs 20 and 21. Supporting characters for each node are given in Supplementary Information.

(Fig. 4a, c). An important question is its orientation. The dorsal side is recognized on the basis of a prominent fin with fin-radials (Fig. 4b). They are inclined towards the preserved end at about 20° to the vertical. In nearly all fish, such an inclination would unequivocally indicate that the posterior end is preserved. There are, however, some rare exceptions to this rule in extant agnathans. In the second dorsal fin of lamprey the fin rays are orientated forwards in the male during the mating period¹⁶. In the hagfish, the rays at the front of the caudal fin may be vertical or even be inclined forwards¹⁷. Moreover, identification of this fossil as the posterior end would be difficult to reconcile with the complex structures, identified with varying degrees of confidence, as part of the branchial basket, cranial cartilages and a pericardic area. To our knowledge, no other living or extinct agnathan has remotely similar structures towards the posterior. The anomalous orientation of these fin-radials, therefore, is assumed to be either comparable to the similar case in the lamprey or alternatively (and less plausibly) due to post-mortem distortion.

In several respects *Haikouichthys* is similar to *Mylokunmingia*. The overall shapes are comparable, although the former is slightly more slender. The presumed head region and trunk are clearly distinguished. In the trunk the myomeres of *Haikouichthys* (Fig. 4a, c) have the same configuration as those of *Mylokunmingia*. The strongly expressed nature of the posteriorly inclined Vs may result from slightly oblique burial. The dorsal fin is more prominent than in *Mylokunmingia*, and differs in possessing closely spaced fin-radials (~7 per mm; Fig. 4b). The rather abrupt scarp between the trunk and ventro-lateral insertion of the fin-fold strongly indicates that originally the fin-folds were paired.

Several structures are visible in the anterior (Fig. 4b, c). Towards the ventral side there is a series of rod-like units, each of which consists of two components: one rod is inclined posteriorly whereas the other is more or less parallel to the ventral margin. Six sets of these units are readily discernible, two more are questionable, and a final set is only tentatively identified (Fig. 4b, c). The total number of sets, therefore, may have been as high as nine. We interpret these rods as part of a branchial basket. No associated gills or filaments are visible. Nearer to the anterior there is a complex series of miner-

alized masses. A preliminary attempt has been made to reconcile some of these areas with both units within the cranial cartilage of the lamprey¹⁸ and the eye. Adjacent to this area, a very short length of the notochord may also be visible. Close to the head–trunk boundary there is an approximately triangular area with mineralized patches. This may represent either the pericardic cartilage or cavity. This, however, assumes that the structures immediately to the posterior are not part of the branchial basket. If these are branchial structures then to the posterior a rather faint pale area, consisting of two regions, is an alternative site for the pericardic cavity. Along the ventral side of the trunk there is a metameric (~13) series of roughly circular structures. These may be the remains of the gonads, although a possible alternative is slime-glands comparable to those of the hagfish¹⁸. A faint trace within the trunk may denote the intestine, and more posteriorly a dark strand may represent its continuation.

Preservation and ecology

The co-occurring fauna includes the trilobites *Eoredlichia* and *Yunnanoccephalus*, as well as exceptionally preserved fossils such as the arthropods *Naraoia* and *Waptia*. Each fossil is preserved in lateral view (Figs 2a, 4a). *Mylokunmingia* gen. nov. is almost complete. One consequence of burial is that the tail has been bent steeply into the sediment, so that the posteriormost section is not visible. This configuration, and what appears to be sediment in the pharynx, suggest that the animal was buried alive. *Haikouichthys* gen. nov. lies parallel to the bedding. The presumed posterior is absent (Fig. 4a, c), possibly owing to decay. There has been growth of sulphide minerals, especially in the head region. The extreme rarity (<0.025%) of chordates in the Chengjiang assemblages⁶ may be because they were active swimmers and could avoid being engulfed by benthic sediment flows that are believed to have been the principal method of burial¹⁹.

Phylogeny

To examine the phylogenetic positions of *Mylokunmingia* and *Haikouichthys* we compiled a data set with 16 taxa (14 vertebrates and two outgroups (tunicate, cephalochordates)) and 116 characters (see Supplementary Information). Most of the characters we used were adopted from existing matrices^{20,21}. Our analysis places both Chengjiang taxa firmly in the agnathans, with both more closely related to lampreys than to hagfishes (Fig. 5). The cladogram is fairly robust when a more cautious attitude is adopted for a number of character states that depend either on specific identifications (such as pericardic cavity) or an absence (for example, arcualia) that might in fact be preservational. Given these provisos, then, in detail *Haikouichthys* forms a monophyletic group with *Jamoytius* and lampreys, but the inter-relationships among them are unresolved. *Mylokunmingia*, however, is evidently more primitive and is basal to the clade comprising both the lamprey–*Jamoytius*–*Haikouichthys* and the 'ostracoderm'–gnathostomes branches. The inclusion of the two Chinese fossils in this cladistic analysis is a significant extension of our knowledge of extinct agnathans. It does not, however, substantially change the internal configuration of existing cladograms in early vertebrates^{18,20–22}.

The phylogenetic implications of this discovery are significant. These agnathan vertebrates predate previous records^{11,23} by at least 20 and possibly 50 Myr (see refs 24 and 25 for Cambrian and early Ordovician radiometric time scales). The differences between *Haikouichthys* and *Mylokunmingia* indicate that, by Chengjiang times, there may already have been a diversity of forms. Neither animal shows evidence for unequivocal biomineralization, and this corroborates the hypothesis that the evolution of bone tissue was a relatively late development in the vertebrate lineage^{18,20}, although the euconodont¹², and possibly paraconodont¹⁴, record indicates that chordate biomineralization was achieved in the Cambrian. In both fossils there is evidence that the ventral fin may have been

paired. The theory of lateral fin folds²⁶ has had considerable significance²¹, but of the fossil agnathans only the anaspids¹⁸ and *Jamoytius*²⁷ have such paired fin-folds. The possible occurrence of this condition in these Lower Cambrian agnathans indicates, however, that fin-folds may be a primitive feature within the vertebrates. The occurrence of close-set dorsal fin-radials in *Haikouichthys*, on the other hand, may be a relatively advanced feature.

Wider implications

The record of Cambrian chordates is very sporadic, and also controversial. Fossils generally accepted to be of a grade comparable to the cephalochordates include the Lower Cambrian *Cathaymyrus*⁶ and Middle Cambrian *Pikaia*⁷. The former genus has been synonymized with the co-eval *Yunnanozoon*²⁸, but this is very doubtful. Moreover, the status of *Yunnanozoon* as a chordate²⁹ is also questionable³⁰, and it may be related to *Xidazoon*³¹ and other stem-group deuterostomes.

The discovery of these Lower Cambrian vertebrates has implications for the likely timing of chordate evolution. The occurrence of *Mylokunmingia* and *Haikouichthys* in the Chengjiang Lagerstätte shows that even more primitive hagfish-like vertebrates had almost certainly evolved by the beginning of the Atdabanian Stage of the Lower Cambrian. Such a find may help to elucidate the transition between the cephalochordates and the first vertebrates. The derivation of the first vertebrates from the cephalochordates must have entailed a major reorganization of the body, especially by the first effective expression of neural-crest tissue³². Key functional steps were the elaboration of the nervous system and active pharyngeal ventilation. However, the range of anatomies seen in the Cambrian cephalochordates^{6,7,9,10} indicates that our documentation of this pre-vertebrate evolutionary plexus is far from complete. It is possible that *Pikaia*, until now the cynosure of Cambrian chordates^{7,8}, is peripheral to the lineage leading to the vertebrates. It has several peculiarities. The head consists of two lobes, each of which bears a prominent, slender tentacle, the mycommata have the reverse configuration of that seen in *Mylokunmingia* and *Haikouichthys*, and the notochord stops well short of the anterior.

The major steps in the early evolution of chordates may well have occurred in the late Neoproterozoic. So far, however, no suitable fossil candidates have been identified amongst the Ediacaran assemblages, although preservation in siltstones and sandstones is less conducive to survival of delicate soft-bodied taxa. Our discovery, however, gives no reason to suppose that the origin of vertebrates was hundreds of millions of years earlier³³, and the reliability of the methods used to reach such a conclusion has been questioned elsewhere^{2,34}. □

Methods

Phylogenetic analysis

The analysis was performed using the software package PAUP 3.1 with a data set of 16 taxa and 116 characters (see Supplementary Information). The analysis was rooted on tunicates. All characters were weighted equally. Characters 62, 63 and 95 were ordered. Most parsimonious trees were identified using the branch and bound search algorithm, stepwise addition. The most parsimonious tree has: tree length, 196; consistency index, 0.628; retention index, 0.700.

Received 5 May; accepted 20 August 1999.

1. Conway Morris, S. *The Crucible of Creation: The Burgess Shale and the Rise of Animals* (Cambridge Univ. Press, Cambridge, 1998).
2. Valentine, J. W., Jablonski, D. & Erwin, D. H. Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development* **126**, 851–859 (1999).

3. Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Phil. Trans. R. Soc. Lond. B* **347**, 305–358 (1995).
4. Budd, G. E. in *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) *Syst. Ass. Spec. Vol. 55*, 125–138 (1997).
5. de Rosa, R. *et al.* Hox genes in brachiopods and priapulids and protostome evolution. *Nature* **399**, 772–776 (1999).
6. Shu, D.-G., Conway Morris, S. & Zhang, X.-L. A *Pikaia*-like chordate from the Lower Cambrian of China. *Nature* **384**, 156–157 (1996).
7. Conway Morris, S. in *Atlas of the Burgess Shale* (ed. Conway Morris, S.) 26 (Palaeontological Association, London, 1982).
8. Briggs, D. E. G., Erwin, D. H. & Collier, F. J. *The Fossils of the Burgess Shale* (Smithsonian, Washington, 1994).
9. Simonetta, A. M. & Insom, E. New animals from the Burgess Shale (Middle Cambrian) and their possible significance for understanding of the Bilateria. *Boll. Zool.* **60**, 97–107 (1993).
10. Insom, E., Pucci, A. & Simonetta, A. M. Cambrian Protochordata, their origin and significance. *Boll. Zool.* **62**, 243–252 (1995).
11. Young, G. C., Karatjute-Talimaa, V. N. & Smith, M. M. A possible Late Cambrian vertebrate from Australia. *Nature* **383**, 810–812 (1996).
12. Donoghue, P. C. J., Purnell, M. A. & Aldridge, R. J. Conodont anatomy, chordate phylogeny and vertebrate classification. *Lethaia* **31**, 211–219 (1998).
13. Pridmore, P. A., Barwick, R. E. & Nicoll, R. S. Soft anatomy and the affinities of conodonts. *Lethaia* **29**, 317–328 (1997).
14. Szaniawski, H. & Bengtson, S. Origin of euconodont elements. *J. Paleontol.* **67**, 640–654 (1993).
15. Luo, H. *et al.* New occurrence of the early Cambrian Chengjiang fauna from Haikou, Kunming, Yunnan province. *Acta. Geol. Sin.* **71**, 97–104 (1997).
16. Marinelli, W. & Strenger, A. *Vergleichende Anatomie und Morphologie der Wirbeltiere, 1. Lampetra fluviatilis (L.)* (Franz Deuticke, Vienna, 1954).
17. Marinelli, W. & Strenger, A. *Vergleichende Anatomie und Morphologie der Wirbeltiere, 2. Myxine glutinosa (L.)* (Franz Deuticke, Vienna, 1956).
18. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, Oxford, 1996).
19. Chen, J.-Y. & Zhou, G.-Q. Biology of the Chengjiang fauna. *Bull. Natl Mus. Nat. Sci. Taiwan* **10**, 11–105 (1997).
20. Forey, P. L. Agnathans recent and fossil, and the origin of jawed vertebrates. *Rev. Fish Biol. Fisheries* **5**, 267–303 (1995).
21. Janvier, P. The dawn of the vertebrates: character versus common ascent in the rise of current vertebrate phylogenies. *Palaeontology* **39**, 259–287 (1996).
22. Forey, P. L. & Janvier, P. Evolution of the early vertebrates. *Am. Sci.* **82**, 554–565 (1994).
23. Young, G. C. Ordovician microvertebrate remains from the Amadeus Basin, central Australia. *J. Vert. Paleont.* **17**, 1–25 (1997).
24. Landing, E. *et al.* Duration of the Early Cambrian: U-Pb ages of volcanic ashes from Avalon and Gondwana. *Can. J. Earth Sci.* **35**, 329–338 (1998).
25. Landing, E. *et al.* U-Pb zircon date from Avalonian Cape Breton Island and geochronologic calibration of the Early Ordovician. *Can. J. Earth Sci.* **34**, 724–730 (1997).
26. Goodrich, E. S. Notes on the development, structure, and origin of the median and paired fins of fish. *Q. J. Microsc. Sci.* **50**, 333–376 (1906).
27. Ritchie, A. New evidence on *Jamoytius kerwoodi*, an important ostracoderm from the Silurian of Lanarkshire. *Palaeontology* **11**, 21–39 (1968).
28. Chen, J.-Y. & Li, C. Early Cambrian chordate from Chengjiang, China. *Bull. Natl Mus. Nat. Sci. Taiwan* **10**, 257–273 (1997).
29. Chen, J.-Y. *et al.* A possible early Cambrian chordate. *Nature* **377**, 720–722 (1995).
30. Shu, D.-G., Zhang, X. & Chen, L. Reinterpretation of *Yunnanozoon* as the earliest known hemichordate. *Nature* **380**, 428–430 (1996).
31. Shu, D. *et al.* A pipiscid-like fossil from the Lower Cambrian of South China. *Nature* **400**, 746–749 (1999).
32. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: A new head. *Science* **220**, 268–274 (1983).
33. Bromham, L. *et al.* Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl Acad. Sci. USA* **95**, 12386–12389 (1998).
34. Ayala, F. J., Rzhetsky, A. & Ayala, F. J. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proc. Natl Acad. Sci. USA* **95**, 606–611 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

Supported by the National Natural Science Foundation of China, Ministry of Sciences & Technology of China, National Geographic USA (D.-G.S.; X.-L.Z.; L.C.; J.H.; Y.L.), Royal Society (S.C.M.), Science & Technology Committee of Yunnan Province (H.-L.L.; S.-X.H.; L.-Z.C.), and Chinese Academy of Sciences (M.Z.). We thank R. J. Aldridge, P. Janvier and R. P. S. Jefferies for helpful remarks, and S. J. Last, S. Capon, D. Simons and L.-H. Li for technical assistance.

Correspondence and requests for materials should be addressed to D.-G.S. (e-mail: dgshu@sein.sxgb.com.cn).

Natural engineering principles of electron tunnelling in biological oxidation–reduction

Christopher C. Page, Christopher C. Moser, Xiaoxi Chen & P. Leslie Dutton

Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

We have surveyed proteins with known atomic structure whose function involves electron transfer; in these, electrons can travel up to 14 Å between redox centres through the protein medium. Transfer over longer distances always involves a chain of cofactors. This redox centre proximity alone is sufficient to allow tunnelling of electrons at rates far faster than the substrate redox reactions it supports. Consequently, there has been no necessity for proteins to evolve optimized routes between redox centres. Instead, simple geometry enables rapid tunnelling to high-energy intermediate states. This greatly simplifies any analysis of redox protein mechanisms and challenges the need to postulate mechanisms of superexchange through redox centres or the maintenance of charge neutrality when investigating electron-transfer reactions. Such tunnelling also allows sequential electron transfer in catalytic sites to surmount radical transition states without involving the movement of hydride ions, as is generally assumed. The 14 Å or less spacing of redox centres provides highly robust engineering for electron transfer, and may reflect selection against designs that have proved more vulnerable to mutations during the course of evolution.

Fifty years ago, Michaelis addressed the problem of oxidoreductase protein design¹, noting that proteins must bring redox centres together. He also recognized that electron transfer between redox prosthetic groups occurs one electron at a time, even though the oxidation–reduction of most biological substrates involves pairwise exchange of electrons. In his words¹ “the task of the protein may be to establish geometric configurations between different prosthetic groups and aid in the formation and stabilization of free radicals”. His view was that these radicals would form upon transfer of a single electron in a thermally activated endergonic step; a second, now exergonic electron transfer would then complete the overall two-electron oxidation or reduction of substrate. Since his time, the structural resolution of dozens of oxidoreductases has fuelled the view that the amino-acid medium has been optimized naturally to guide electron tunnelling and finely tune favourable free energies (ΔG) and allied parameters. Moreover, endergonic tunnelling has been effectively ignored in physiological reactions. Here we uncover simple and general guidelines that describe an unusually robust natural engineering quite different from this prevailing viewpoint.

Electron-transfer rate calculation

A general view of electron-transfer engineering requires an examination of a broad range of oxidoreductases and a direct means of calculating the rates at which electrons tunnel through insulating protein. This tunnelling rate is commonly viewed as the product of two terms². The first is an electronic term arising from the strength of the coupling of the electron donor/acceptor wavefunctions, leading to roughly exponential fall-off in the electron tunnelling rate with distance through the insulating barrier and is proportional to $e^{-\beta R}$, where R is the edge-to-edge distance and β is proportional to the square root of the barrier height. The second term depends on the energy, λ , required to reorganize nuclear coordinates upon electron transfer, and the driving force, ΔG , for the electron transfer. Both classical³ and quantum^{4,5} versions of Marcus electron transfer theory suggest a roughly parabolic dependence of log rate on ΔG (inset, Fig. 1d). The electronic terms of different reactions can be compared by using the ΔG optimized rate ($\Delta G_{\text{opt}} = -\lambda$)⁶.

The overall approximately exponential fall-off of the ΔG optimized

rate with tunnelling distance is known to be significantly modulated by the intervening medium. Electron tunnelling between covalently bridged redox centres in synthetic systems ($\beta \approx 0.9 \text{ Å}^{-1}$)⁷ is clearly much faster than tunnelling through vacuum (β varies from 2.8 Å^{-1} (ref. 6) to 3.5 Å^{-1} (ref. 8)). Our earlier experimental examination of tunnelling in proteins suggested an intermediate value ($\beta = 1.4 \pm 0.2 \text{ Å}^{-1}$) corresponding to a weighted average of the two extreme β values⁶. A simple empirical expression that incorporates an exponential decay of tunnelling rate k_{et} (in s^{-1}) with edge-to-edge distance R (in Å) and a parabolic dependence of log rate on ΔG and λ (in eV) is given by equation (1) (ref. 9).

$$\log_{10} k_{\text{et}} = 15 - 0.6 R - 3.1(\Delta G + \lambda)^2 / \lambda \quad (1)$$

The coefficient 0.6 reflects $\beta = 1.4 \text{ Å}^{-1}$ on a common log scale, whereas the coefficient 3.1 collects the room temperature constants for the quantized nuclear term⁵, as suggested by extensive studies in photosynthetic reaction centres^{6,10,11}. This equation has proven useful as a first approximation in the absence of a detailed structure; however, it does not explicitly address the variations in polypeptide structure, nor whether those variations have been naturally selected to influence tunnelling rates to physiological advantage.

Equation (2) gives an empirical expression that includes the packing density (ρ) of protein atoms in the volume between redox centres to account for the β variations in an exergonic electron tunnelling rate $k_{\text{et}}^{\text{ex}}$.

$$\log_{10} k_{\text{et}}^{\text{ex}} = 13.0 - (1.2 - 0.8\rho)(R - 3.6) - 3.1(\Delta G + \lambda)^2 / \lambda \quad (2)$$

We define ρ as the fraction of the volume between redox cofactors that is within the united van der Waals radius of intervening atoms. In principle, ρ can range from 1, corresponding to a fully packed medium ($\beta = 0.9 \text{ Å}^{-1}$), to a value of 0, corresponding to the interstitial space in the protein structure outside the united van der Waals atomic radii ($\beta = 2.8 \text{ Å}^{-1}$). The ρ weighting of these betas, $\beta = (\rho)0.9 \text{ Å}^{-1} + (1 - \rho)2.8 \text{ Å}^{-1}$, generates the $(1.2 - 0.8\rho)$ coefficient in equation (2). The log of the optimal rate at 3.6 Å van der Waals contact is 13 (ref. 6). Figure 1a compares this elementary analysis with measured tunnelling rates in two photosynthetic reaction centre proteins, and indicates a remarkably small threefold standard deviation.

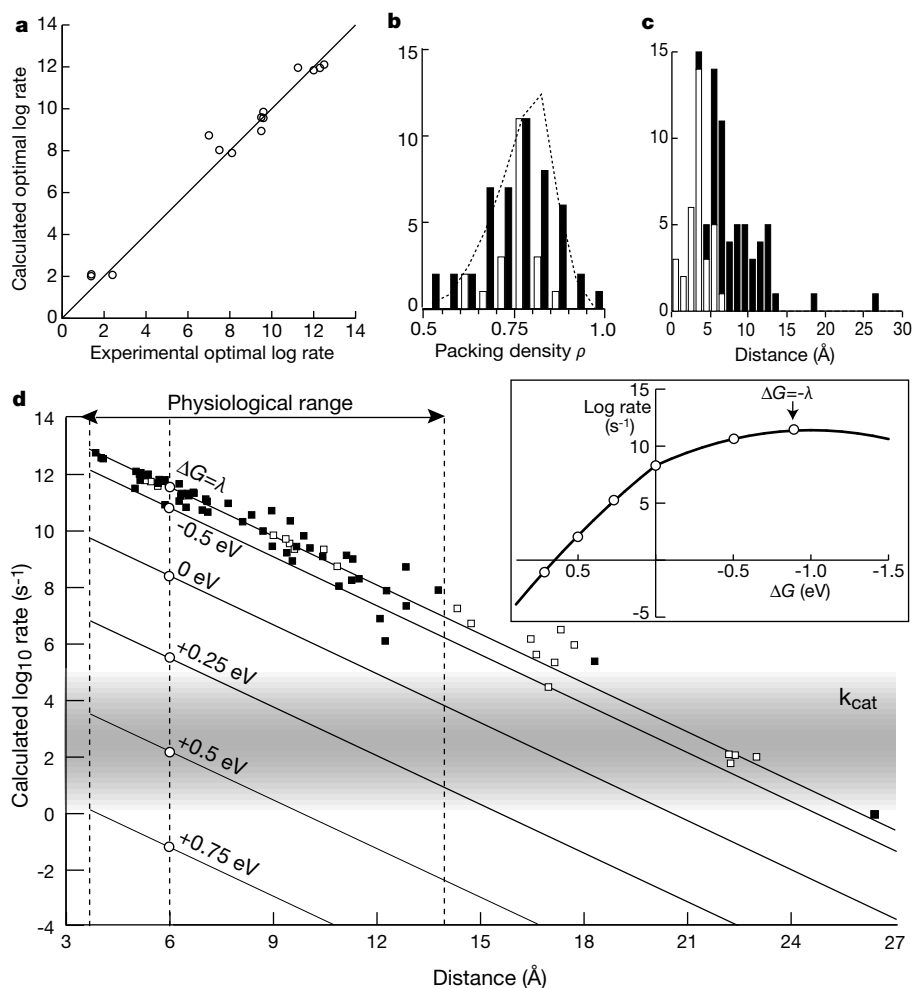


Figure 1 Electron-transfer rates and statistics for natural multiredox centre oxidoreductase structures in the PDB as of February 1999. **a**, Comparison of calculated (equation (2)) with experimental free-energy-optimized electron-transfer rates for the photosynthetic reaction centres of *Rhodobacter sphaeroides* and *Rp. viridis*. The standard deviation is 0.5 log units. Measured rates, ΔG and λ values are taken from refs 6, 23, 45–47. **b**, Distributions of packing density ρ for physiologically productive (solid bars) and unphysiological (open bars) electron-transfer reactions for proteins in the PDB database. The dashed line shows a normalized distribution of the presumably arbitrary interior packing found between standard donor cofactors and the 6-atom aromatic ring of tyrosine and phenylalanine side chains as mock cofactors. **c**, A histogram of edge-to-edge distances of long-range physiological electron transfer for 31 different redox proteins in the PDB database shows that almost all productive reactions are less than 14 Å. The two examples at longer distances may represent physiologically inappropriate structures (see

text). Solid bars, distances between cofactors within chains; open bars, distances between closely spaced substrates and redox centers within catalytic clusters. **d**, Free-energy-optimized rate of electron tunnelling, predicted by equation (2), for the electron-transfer proteins of **b** and **c**, excluding for clarity substrates and those redox centres in van der Waals or bonded contact. Physiologically productive reactions (solid squares) and unproductive reactions (open squares) fall evenly on either side of an average log rate versus distance relationship ($\Delta G = -\lambda$). The calculated rates for non-optimal ΔG can be gauged from the parallel lines, which use an unremarkable $\lambda = 1$ eV. Inset, the expected free-energy dependence of the rate for both exergonic and endergonic tunnelling reactions for a fixed distance of 6 Å and the same generic λ . Circles correspond to free energies chosen for the diagonal lines shown in the main figure. The grey zone gives the typical range of turnover rates (k_{cat}) of electron-transfer proteins.

Other tunnelling analyses also reveal protein structural heterogeneity, including methods that search for the shortest relatively well-bonded pathway between donor and acceptor⁸ or the Monte Carlo scattering method¹². But given the current state of experimental uncertainty arising from structural changes due to protein dynamics and unidentified solvent molecules, as well as significant errors associated with estimating the ΔG optimized rates, no one method has been clearly shown to be superior. Until experimental parameters are better defined, the simple and easily understood packing approach provides the necessary accuracy.

Survey of electron-transfer proteins

We have examined oxidoreductase proteins that have defined function, multiple cofactors for intramolecular electron transfer and a structure deposited in the Protein Data Bank (PDB). Figure 1b shows that the packing fraction of the protein, ρ , is statistically

identical whether measured between redox centres of physiologically productive reactions ($\rho = 0.76 \pm 0.10$; $\beta = 1.39 \pm 0.15$), between redox centres of unphysiological, potentially counterproductive electron transfers ($\rho = 0.76 \pm 0.06$), or in the presumably arbitrary regions between redox centres and aromatic side chains scattered throughout the protein ($\rho = 0.77 \pm 0.09$). These results show that nature does not use protein structural heterogeneity to physiological benefit.

In an alternative measure of the effect of protein structural heterogeneity, we find that well-bonded pathways radiate from a given redox centre in many different directions. We also find that protein regions with well-bonded pathways tend to be well-packed. As in the packing analysis, there does not appear to be any general correlation between the quality of the path and the physiological benefit of the reaction.

Figure 1c shows that physiological distances between the edges of

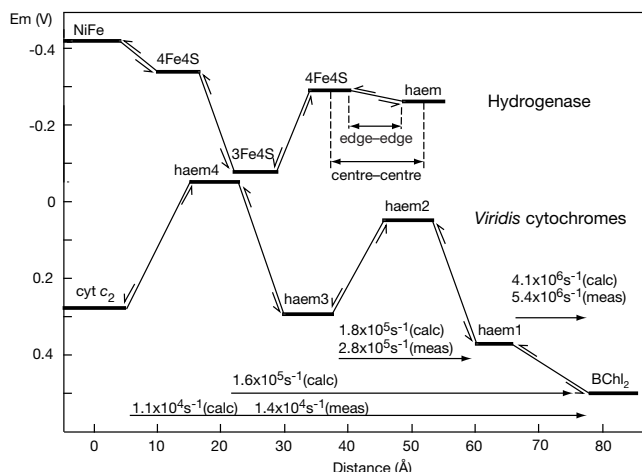


Figure 2 Chains of redox centres are used to cover large distances in natural proteins; chains often include conspicuously exergonic and endergonic electron-transfer steps. The two examples shown are the haem chain from cytochromes of *Rp. viridis* and the FeS cluster chain from hydrogenase of *Ds. gigas*. The spacing between horizontal bars representing redox cofactors reflects the edge-to-edge distance, whereas the centre-to-

centre distance is measured from the centre of each bar (from PDB entries 1prc, 1frv, and ref. 48). The *y* axis indicates the reported redox midpoint potential (E_m) of each centre in the chains^{22,49,50}. The several measured rates in *Rp. viridis*^{22,25} are compared with calculated rates using a uniform reorganization energy of 0.7 eV.

redox centres are almost all within 14 Å. In the two exceptions, interdomain flexibility in sulphite oxidase may shorten the ~27 Å between haem *b* and the molybdenum centre¹³, and physiological electron transfer between non-haem iron centres separated by ~18 Å in desulphoferrodoxin has not been shown¹⁴. Figure 1d shows that within the 14 Å tunnelling boundary, the calculated ΔG -optimized tunnelling rates are very high: in the 10^{13} – 10^7 s^{−1} range. Figure 1d also shows that tunnelling rates remain higher than catalytic rates (k_{cat}) at typical physiological ΔG (0.0 to −0.1 eV). Moreover, at the shorter distances, strikingly uphill tunnelling can occur at functionally relevant rates. For example, at the 6 Å distance marked in Fig. 1d, and with unremarkable values of λ and ρ , the tunnelling rate is maximal at 3×10^{11} s^{−1} when ΔG matches $-\lambda$, remains high at 2×10^8 s^{−1} when $\Delta G = 0$, and is still 10^2 s^{−1} for a +0.5 eV endergonic step.

Thus, we find that by favouring proteins with redox centres placed within 14 Å of each other, natural selection fosters robust electron-transfer designs. The designs provide very broad engineering tolerances to changes in free energy, including endergonic tunnelling; they are also relatively insensitive to changes in the detailed protein structure and even to changes in the chemical nature of the redox centres themselves⁶. The enormous latitude in values of ρ , λ and ΔG , which sustain tunnelling rates faster than typical catalytic rates, shrinks sharply at distances greater than 14 Å where designs become less robust.

Redox chains for long-range tunnelling

Natural systems often need to transfer electrons with high directional specificity over distances greater than 14 Å. This is routinely accomplished by multistep tunnelling through chains of redox centres individually spaced within the 14 Å limit. Although the overall reactions are usually modestly favourable, these chains often include very endergonic steps¹⁵. There is a frequent reluctance to consider endergonic tunnelling in physiological reactions; however, tunnelling up to 0.48 eV uphill has been quantitatively shown by systematic chemical modification of photosynthetic reaction centres^{16,17}. Equation (3) estimates the rate of endergonic tunnelling steps by using equation (2) for the opposite exergonic step and dividing by the temperature-dependent Boltzmann factor, or equilibrium constant ($10^{\Delta G/0.06}$ at room temperature).

$$\log_{10} k_{et}^{en} = 13.0 - (1.2 - 0.8\rho)(R - 3.6) - 3.1(-\Delta G + \lambda)^2/\lambda - \Delta G/0.06 \quad (3)$$

In the iron sulphur clusters/cytochrome *c*₃ chain of *Desulfovibrio gigas* Ni–Fe hydrogenase (Fig. 2)¹⁸, calculations reveal that rates of transit over tens of ångströms are in the 2×10^7 s^{−1} range (using a λ of ~0.7 eV), some three or four orders of magnitude faster than the overall k_{cat} , despite a 0.21 eV endergonic step. *Desulfovibrio fructosovorans* has an even larger 0.40 eV step¹⁹, for which we calculate a rate of 5×10^4 s^{−1}. Figure 2 also shows the haem chain of the *Rhodospseudomonas viridis* reaction centre^{20,21}. The agreement between our calculated rates with the several measured rates in individual segments^{22,23} and the whole chain^{24,25} further supports the utility of equation (3) and shows that endergonic tunnelling can be a normal part of physiological electron-transfer chains.

Modelling of these systems shows that natural proximity, together with thermal activation, provides electron-transfer chains with a robustness to substantial variance in ΔG , λ and ρ , before k_{cat} becomes limited by tunnelling. Thus, a redox centre should not be prematurely rejected from a chain simply because its midpoint potential is unfavourable, nor should the centre automatically be relegated to a virtual role in the distinctly different mechanism of redox-centre superexchange. Proximity of redox centres in chains provides rugged, highly directional electron transfer, which slows roughly linearly with distance along the chain, in contrast to the sharply exponential slowing in all other directions.

Design of catalytic clusters

Outside catalytic sites, organic substrates are resistant to oxidation and reduction because single electron transfer to form a radical intermediate state can rarely be followed by a rapid second electron transfer to form a stable redox state. Catalysis is achieved in proteins through the design of redox clusters in which at least two electron donors or acceptors are secured in close proximity to the substrate (≤ 6 Å in Fig. 1c). At these distances, catalytic sites can promote endergonic tunnelling to levels even higher than those described for chains. Equations (2) and (3) can again be used to explore and quantify substrate oxidation–reduction.

Figure 3a shows cytochrome *c* oxidase with two one-electron redox centres—haem *a*₃ and Cu_B—flanking substrate O₂, the reduction of which we follow to just the two-electron peroxy stage. Thermally assisted endergonic tunnelling to the superoxide radical state ($\Delta G = 0.58$ eV, $k_{stab} \approx 10^{-20}$) is followed by a fast exergonic second-electron transfer. Rate calculations in this model accommodate the measured rate of 5×10^3 s^{−1} (ref. 26) and an overall modest ΔG of −0.12 eV (ref. 27). The redox potential of the

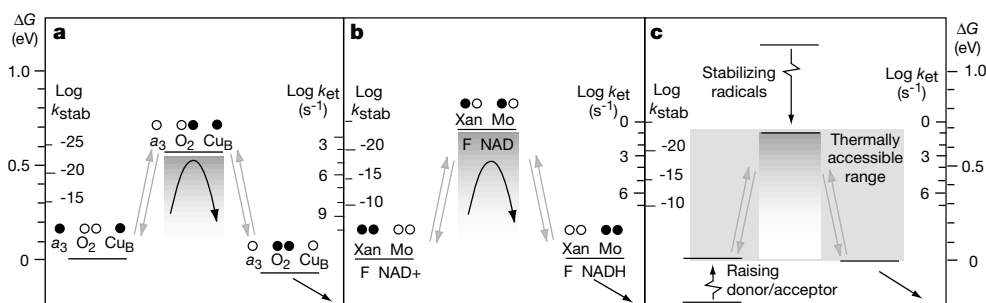


Figure 3 The proximity between redox centres in clusters allows rapid tunnelling access to considerably endergonic radical intermediate states. **a**, A three-member cluster in cytochrome oxidase: haem a_3 , oxygen and Cu_B . Haem a_3 and Cu_B are reduced (filled circles) at the start. The overall rate of electron transfer to doubly reduced oxygen, given by equations (2) and (3) ($\log_{10} k_{et}$), will depend on the stability constant for the O_2^- superoxide radical state ($\log_{10} k_{stab}$). An endergonic step over a range of up to +0.6 eV

oxygen/superoxide radical in this calculation is close to that in neutral aqueous solution (−0.2 V)²⁷, which indicates that catalysis of dioxygen reduction can proceed even without stabilization of the superoxide radical.

Figure 3b shows the analogous schematic for a cluster including a two-electron redox centre adjacent to a substrate. Xanthine dehydrogenase provides two examples. This enzyme is a typical flavin-linked nicotinamide adenine dinucleotide (NAD)(P)H oxidoreductase in which flavin provides two single-electron redox transitions, with potentials separated by 0.14 V ($k_{stab} \approx 10^2$) and a small overall free energy for oxidation or reduction of the nicotinamide substrate²⁸. The close proximity of flavin and NAD enables endergonic electron tunnelling of more than 0.7 eV (to a potential of −1.1 V) to form the NAD $\cdot$ radical, followed by an exergonic second transfer to form the product well within the 1–10 s^{−1} k_{cat} range typical of these enzymes. Again, two single-electron transfers can be easily accomplished without reliance on the site modifying the potential of the aqueous NAD $\cdot$ /NAD $^+$ redox couple of −0.92 V (ref. 29). Xanthine dehydrogenase includes another cluster, akin to the molybdopterine centre in aldehyde dehydrogenase, in which xanthine can reasonably be oxidized to its acid through two single-electron steps reducing Mo^{VI} to Mo^V and Mo^V to Mo^{IV} , with an overall ΔG that is small and favourable^{30,31}. Although the level of the postulated radical– Mo^V transition state is not known, our calculations show that it can be as high as 0.8 eV (18.5 kcal mol^{−1}) and still maintain a tunnelling rate faster than the observed k_{cat} of 1–10 s^{−1}.

Sequential electron transfer rates can exceed k_{cat} even when the overall rate is slowed by an associated H $^+$ or H $\cdot$ radical movement, provided that redox centres are closely clustered. Isotope sensitivity does not exclude a sequential single-electron transfer mechanism³².

The sequential, one-electron transfer Michaelis-like view that we describe here opposes the traditional concerted two-electron model for these reactions, in the way that sequential single-electron transfer radical transition state mechanisms oppose traditional two-electron mechanisms in solution organic chemistry^{33,34}. Abundant examples of tunnelling in solution to the radical transition state provide Marcus plots highly similar to the inset of Fig. 1d, and include measurements that support the sequential electron-transfer mechanism well into the endergonic range. If substrate redox catalysis in biology follows the principles learned from solution studies, then there will be many more examples of sequential electron transfer/radical transition state mechanisms than have been considered to date.

Design and evolution

Although it is commonly believed that the detailed structure of oxidoreductases has been finely tuned and optimized for electron

allows electron tunnelling to be faster than the experimental k_{obs} of 5×10^3 s^{−1}.

b, Corresponding examples of two-centre clusters of flavin (F) and NADH in xanthine dehydrogenase as well as the xanthine (Xan) molybdenum (Mo) pair. **c**, Examples of general strategies in which radical states are moved into the thermally accessible regimen by stabilizing the radical intermediate at the catalytic site, or raising the levels of the initial donor or acceptor closer to the radical state.

transfer by millions of years of natural selection, it is clear from our analysis that these proteins are relatively ruggedly built so that individual electron transfer rates do not have to be optimized. The single characteristic that has been obviously selected in the evolution of physiological electron tunnelling in oxidoreductases is the simple positioning of redox centres within 14 Å of each other. The high rates and directional specificity achieved by proximity relieves evolutionary pressure to adjust polypeptide structures between redox centres to influence tunnelling. Accordingly, nature does not appear to construct a tightly packed protein medium with well-bonded pathways between natural redox partners to enhance productive electron transfers, or to construct loosely packed or poorly bonded regions to guard against adverse reactions. Nor have redox centres been positioned by natural selection to exploit potential packing differences between α -helices and β -sheets³⁵.

The general robustness of oxidoreductases to amino-acid substitution has been shown by many mutational studies, such as the structural and kinetic studies of the systematic replacement of Tyr 162 that lies directly between the haem and bacteriochlorophyll dimer in the *Rp. viridis* photosynthetic reaction centre²³. Mutations frequently produce modest effects on the tunnelling rate (in this example, up to fivefold) that can be understood in terms of equation (2) with variations in R (± 1 Å) and ΔG (± 0.1 V) (ref. 23) and ρ (± 0.2). Despite the mutational evidence, there is a persistent practice of tracking ‘best pathways’ in protein structures as though particular bonds or amino acids that happen to lie between redox centres are examples of natural engineering to guide the tunnelling electron. This tracking disregards that almost every change of an amino acid will provide alternative bonds to form a comparable pathway. Another common pathway misconception is that water molecules between redox centres cannot support electron tunnelling because they are not linked to the protein. A packing perspective, which emphasizes the properties of the intervening medium as a whole, avoids these misconceptions and indicates that amino acids and structural motifs in these regions have many other biologically important functions with their own selective pressures not directly related to electron tunnelling. These multiple roles which dampen natural selection have been recognized³⁶, and hark back to Darwin’s own description of natural selection³⁷.

In chains, the high tunnelling rates provided by redox centre proximity also relieve evolutionary pressure to finely tune free energies and reorganization energies. Indeed, as distances are shortened from the 14 Å boundary, increasingly endergonic electron tunnelling steps can still support rapid electron transfer. Thus, electron-transfer states that might normally be dismissed because they are energetically unfavourable by hundreds of meV can become reasonable reaction intermediates, although they may be present at

extremely small concentrations. These thermally accessed states are real intermediates, not the virtual intermediates of superexchange, whose function in room-temperature biology is far from established. For an alternative view see ref. 38.

The relaxed demands on ΔG for intermediate steps can translate into a substantial toleration of slow or incomplete charge compensation during oxidation–reduction. For many chains, the only engineering of ΔG that may be required is a rather coarse charge compensation and proton management in the form of a distribution of acid and base side chains near the chain redox centres. More refined and mutagenically far more fragile structures, such as the extensive 35 Å fully charge-compensated hydrogen-atom transfer pathway hypothesized for ribonucleotide reductase^{39,40}, are conceivable but may not be necessary. Our calculations show that the chain of amino-acid radicals in ribonucleotide reductase can tolerate endergonic steps (perhaps owing to imperfect charge compensation) of up to 0.45 eV before an effect on k_{cat} is observed. As in many chains, electron transfer may take place without precise choreography of H^+ movement to assure charge neutrality; H^+ release and uptake may be coupled only at the origin of the chain and at the substrate site.

The simplest oxidoreductases are cofactor chains. Natural design places redox partners close enough to outpace competing reactions. For example, light-energy conversion in photosynthetic reaction centres requires little more than rapid charge separation to defeat nanosecond decay processes such as fluorescence. This is achieved by spacing chlorin centres close together to create two rapid steps, each driven by small favourable ΔG (see refs. 6,9).

In many substrate-binding oxidoreductases, this same proximity allows two separate electron-transfer reactions to be combined with thermal activation to surmount substrate radical transition-state levels. Such a mechanism, unequivocal in the bifurcated electron transfer from ubihydroquinone in the cytochrome bc_1 complex⁴¹, is readily adapted for many other substrates beyond the O_2 and NAD described here. In many cases, little or no electrochemical stabilization of the bound substrate radical transition states may be required.

However, the radical redox levels of some free substrates may be outside the limits of our examples. In these cases, specific engineering may be required to lower the height of the endergonic tunnelling step. Figure 3c shows two methods: a traditional stabilization of the radical transition state of bound substrate, akin to the action of classical enzymes; and movement of the donor or acceptor redox levels towards the substrate radical level. For example, ATP hydrolysis substantially lowers the redox potentials of donors in nitrogenase, moving them towards the super-reduced molybdenum level⁴², which we propose would be within the endergonic tunnelling range of the levels of the nitrogen radical states. In a complementary example, remotely generated side-chain radicals are relayed to the catalytic site of ribonucleotide reductase, markedly raising the redox levels of the tyrosine acceptor towards the ribose ring radical states and bringing the sequential two-electron ribose ring reduction within range⁴³. Only when the simple robust designs of single-electron endergonic tunnelling are exhausted may nature select more sophisticated, geometrically constrained site designs for single-step two-electron mechanisms of hydride transfer⁴⁴. □

Methods

Defining R and cofactors

Edge-to-edge distance was defined as the minimum distance between the coordinates of a donor cofactor atom and the coordinates of an acceptor cofactor atom. Cofactors were defined as all redox metal atoms, plus all atoms directly liganded to the metal atoms. The associated aromatic rings of flavins, quinones, radicals, pterins and porphyrins were also included, as were the oxygen atoms attached to tyrosine or the quinone aromatic rings.

Determining ρ

Each structure was solvated with water by the droplet method using the molecular modelling program Sybyl (Tripos Associates). We assayed protein heterogeneity by a

simple sampling of the density of protein packing in the volume between redox centres. Larger volumes, such as the roughly ellipsoidal volumes previously described¹², have progressively less significance for tunnelling. ρ was calculated by drawing a line from each atom in the donor cofactor to each atom in the acceptor cofactor. Each line was binned in 0.2 Å segments. The packing, ρ , was the fraction of those bins within the united atom van der Waals radius of a non-hydrogen atom, excluding from the calculation those bins overlapping a cofactor atom.

For distances under 5 Å, this method of estimating ρ picks up a large statistical noise due to very few non-cofactor bins; however, at these short distances a large variability in ρ has little impact on the rate (see Fig. 1d). The distributions in Fig. 1b exclude those high noise points for clarity.

Protein data set

Our data set consisted of all natural multiredox centre oxidoreductase structures in the PDB as of February 1999. The physiological relevance of the small number of intermolecular structures is too uncertain to warrant inclusion in this survey.

The proteins performing the physiological reactions were aldehyde ferredoxin oxidoreductase (laor), aldehyde oxidoreductase (lalo), ascorbate oxidase (laox), copper amine oxidase (lksi), cytochrome bc_1 complex (3bcc), cytochrome c oxidase (1arl), cytochrome c peroxidase (2pcc), cytochrome c_3 (2cym), cytochrome c_4 (1etp), cytochrome cd /nitrite reductase (1aof), desulphoferredoxin (1dfx), ferredoxin (1fdx and 7ddl), flavocytochrome b_2 (1fcb), flavocytochrome c /sulphide dehydrogenase (1fcd), formate dehydrogenase (1fdi), hydroxylamine oxidoreductase (1fgh), iron-only hydrogenase (1feh), lignin peroxidase (1b80), NADH/cytochrome $P450$ reductase (1amo), Ni–Fe hydrogenase (1frv), nitrite reductase (1nid), nitrogenase (1n2c), peptidylglycine α -hydroxylating monooxygenase (1phm), photosynthetic reaction centre from *Rp. viridis* (1prc), phthalate dioxygenase reductase (2pia), prostaglandin H_2 synthase (1prh), pyruvate ferredoxin oxidoreductase (1b0p), quinone reductase (1qrd), ribonucleotide reductase R2 subunit (1rib), sulphite oxidase (1sox), sulphite reductase (1aop) and trimethylamine dehydrogenase (2tmd). Non-physiological reactions included photosynthetic reaction centres (1prc and 1pcr, except for reactions involving Q_B where 4rcr was more appropriate) and cytochrome oxidases (1occ and 1arl); as these reactions were relatively few in number, two species of each type were used to improve the unproductive distribution. The proteins used for examining the arbitrary interior packing in Fig. 1b were nitrogenase, Ni–Fe hydrogenase, cytochrome c oxidase, flavocytochrome b_2 , the h and l subunits of methylamine dehydrogenase (2mta) and copper amine oxidase (1oac).

Received 23 July; accepted 16 September 1999.

1. Michaelis, L. in *The Enzymes, Chemistry and Mechanism of Action* (eds Sumner, J. B. & Myrback, K.) 1–54 (Academic, New York, 1951).
2. Devault, D. Quantum mechanical tunnelling in biological systems. *Quart. Rev. Biophys.* **13**, 387–564 (1980).
3. Marcus, R. A. & Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta* **811**, 265–322 (1985).
4. Jortner, J. Temperature dependent activation energy for electron transfer between biological molecules. *J. Chem. Phys.* **64**, 4860–4867 (1976).
5. Hopfield, J. J. Electron transfer between biological molecules by thermally activated tunneling. *Proc. Natl Acad. Sci. USA* **71**, 3640–3644 (1974).
6. Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S. & Dutton, P. L. Nature of biological electron-transfer. *Nature* **355**, 796–802 (1992).
7. Smalley, J. F. et al. The kinetics of electron-transfer through ferrocene-terminated alkanethiol monolayers on gold. *J. Phys. Chem.* **99**, 13141–13149 (1995).
8. Beratan, D. N., Onuchic, J. N., Winkler, J. R. & Gray, H. B. Electron-tunneling pathways in proteins. *Science* **258**, 1740–1741 (1992).
9. Moser, C. C. & Dutton, P. L. Engineering protein structure for electron transfer function in photosynthetic reaction centers. *Biochim. Biophys. Acta* **1101**, 171–176 (1992).
10. Gunner, M. R. & Dutton, P. L. Temperature and ΔG° dependence of the electron-transfer from BPh^+ to Q_A in reaction center protein from *Rhodospirillum rubrum* with different quinones as Q_A . *J. Am. Chem. Soc.* **111**, 3400–3412 (1989).
11. Venturoli, G. et al. Effects of temperature and ΔG° on electron transfer from cytochrome c_2 to the photosynthetic reaction center of the purple bacterium *Rhodospirillum rubrum*. *Biophys. J.* **74**, 3226–3240 (1998).
12. Kuki, A. & Wolynes, P. G. Electron tunneling paths in proteins. *Science* **236**, 1647–1652 (1987).
13. Kisker, C. et al. Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell* **91**, 973–983 (1997).
14. Coelho, A. V. et al. Desulphoferredoxin structure determined by MAD phasing and refinement to 1.9-angstrom resolution reveals a unique combination of a tetrahedral FeS_4 centre with a square pyramidal FeS_4 centre. *J. Biol. Inorg. Chem.* **2**, 680–689 (1997).
15. Dutton, P. L. et al. in *Biological Electron Transfer Chains: Genetics, Composition and Mode of Operation* 3–8 (Kluwer Academic, Netherlands, 1998).
16. Gunner, M. R., Tiede, D. M., Prince, R. C. & Dutton, P. L. in *Function of Quinones in Energy Conserving Systems* (ed. Trumpower, B. L.) 265–269 (Academic, New York, 1982).
17. Woodbury, N. W., Parson, W. W., Gunner, M. R., Prince, R. C. & Dutton, P. L. Radical-pair energetics and decay mechanisms in reaction centers containing anthraquinones, naphthoquinones or benzoquinones in place of ubiquinone. *Biochim. Biophys. Acta* **851**, 6–22 (1986).
18. Volbeda, A. et al. Crystal-structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).
19. Roussel, M. et al. [3Fe-4S] to [4Fe-4S] cluster conversion in *Desulfovibrio fructosovorans* [NiFe] hydrogenase by site-directed mutagenesis. *Proc. Natl Acad. Sci. USA* **95**, 11625–11630 (1998).
20. Michel, H., Deisenhofer, J. & Epp, O. Pigment protein interactions in the photosynthetic reaction center from *Rhodospirillum rubrum*. *EMBO J.* **5**, 2445–2451 (1986).
21. Knaff, D. B. et al. Reaction of cytochrome c_2 with photosynthetic reaction centers from *Rhodospirillum rubrum*. *Biochemistry* **30**, 1303–13010 (1991).

22. Shopes, R. J., Holten, D., Levine, L. M. A. & Wraight, C. A. Kinetics of oxidation of the bound cytochromes in reaction centers from *Rhodospseudomonas viridis*. *Photosynth. Res.* **12**, 165–180 (1987).
23. Dohse, B. *et al.* Electron-transfer from the tetraheme cytochrome to the special pair in the *Rhodospseudomonas viridis* reaction-center—effect of mutations of tyrosine L162. *Biochemistry* **34**, 11335–11343 (1995).
24. Meyer, T. E., Bartsch, R. G., Cusanovich, M. A. & Tollin, G. Kinetics of photooxidation of soluble cytochromes, h₁ and h₂, and azurin by the photosynthetic reaction center of the purple phototrophic bacterium *Rhodospseudomonas viridis*. *Biochemistry* **32**, 4719–4726 (1993).
25. Ortega, J. M., Drepper, F. & Mathis, P. Electron transfer between cytochrome c₂ and the tetraheme cytochrome c in *Rhodospseudomonas viridis*. *Photosyn. Res.* **59**, 147–157 (1999).
26. Proshlyakov, D. A., Pressler, M. A. & Babcock, G. T. Dioxygen activation and bond cleavage by mixed-valence cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **95**, 8020–8025 (1998).
27. Stubbe, J. & van der Donk, W. A. Protein radicals in enzyme catalysis. *Chem. Rev.* **98**, 705–762 (1998).
28. Hunt, J., Massey, V., Dunham, W. R. & Sands, R. H. Redox potentials of milk xanthine dehydrogenase—room-temperature measurement of the FAD and 2Fe/2S center potentials. *J. Biol. Chem.* **268**, 18685–18691 (1993).
29. Farrington, J. A., Land, E. J. & Swallow, A. J. The one-electron reduction potentials of NAD. *Biochim. Biophys. Acta* **590**, 273–276 (1980).
30. Romao, M. J. *et al.* Crystal-structure of the xanthine oxidase-related aldehyde oxidoreductase from *D. gigas*. *Science* **270**, 1170–1176 (1995).
31. Huber, R. *et al.* A structure-based catalytic mechanism for the xanthine oxidase family of molybdenum enzymes. *Proc. Natl Acad. Sci. USA* **93**, 8846–8851 (1996).
32. Powell, M. F., Wu, J. C. & Bruice, T. C. Ferricyanide oxidation of dihydropyridines and analogues. *J. Am. Chem. Soc.* **106**, 3850–3856 (1984).
33. Ebersson, L. *Electron Transfer Reactions in Organic Chemistry* 1–234 (Springer, New York, 1987).
34. Pross, A. The single electron shift as a fundamental process in organic chemistry: the relationship between polar and electron-transfer pathways. *Acc. Chem. Res.* **18**, 212–219 (1985).
35. Gray, H. B. & Winkler, J. R. Electron transfer in proteins. *Annu. Rev. Biochem.* **65**, 537–561 (1996).
36. McLendon, G. Control of biological electron-transport via molecular recognition and binding—the velcro model. *Struct. Bond.* **75**, 159–174 (1991).
37. Darwin, C. *Origin of Species by Means of Natural Selection* (Earlton House, New York, 1872).
38. Jortner, J., Bixon, M., Langenbacher, T. & Michel-Beyerle, M. E. Charge transfer and transport in DNA. *Proc. Natl Acad. Sci. USA* **95**, 12759–12765 (1998).
39. Siegbahn, P. E. M., Blomberg, M. R. A. & Crabtree, R. H. Hydrogen transfer in the presence of amino acid radicals. *Theor. Chem. Acc.* **97**, 289–300 (1997).
40. Ehrenberg, A. Protein dynamics, free radical transfer and reaction cycle of ribonucleotide reductase. *Third International Symposium on Biological Physics* **1**, 1–10 (Springer, New York, 1998).
41. Mitchell, P. The protonmotive Q cycle: a general formulation. *FEBS Lett.* **59**, 137–139 (1975).
42. Spee, J. H. *et al.* Redox properties and electron paramagnetic resonance spectroscopy of the transition state complex of *Azotobacter vinelandii* nitrogenase. *FEBS Lett.* **432**, 55–58 (1998).
43. Siegbahn, P. E. M. Theoretical study of the substrate mechanism of ribonucleotide reductase. *J. Am. Chem. Soc.* **120**, 8417–8439 (1998).
44. Bahnsen, B. J., Colby, T. D., Chin, J. K., Goldstein, B. M. & Klinman, J. P. A link between protein structure and enzyme catalyzed hydrogen tunneling. *Proc. Natl Acad. Sci. USA* **94**, 12797–12802 (1997).
45. Zinth, W., Arlt, T. & Wachtveitl, J. The primary processes of bacterial photosynthesis—ultrafast reactions for the optimum use of light energy. *Ber. Bunsenges Phys. Chem.* **100**, 1962–1966 (1996).
46. Allen, J. P. *et al.* Free energy dependence of the direct charge recombination from the primary and secondary quinones in reaction centers from *Rhodobacter sphaeroides*. *Photosyn. Res.* **55**, 227–233 (1998).
47. Li, J. L., Gilroy, D., Tiede, D. M. & Gunner, M. R. Kinetic phases in the electron transfer from P⁺Q_A–Q_B to P⁺Q_AQ_B– and the associated processes in *Rhodobacter sphaeroides* R-26 reaction centers. *Biochemistry* **37**, 2818–2829 (1998).
48. Sogabe, S. *et al.* Crystal-structure of cytochrome-c₂ from *Rhodospseudomonas viridis* at 3.0 Å resolution. *Photosyn. Res.* **34**, 154–154 (1992).
49. Alegria, G. & Dutton, P. L. Langmuir–Blodgett monolayer films of bacterial photosynthetic membranes and isolated reaction centers: preparation, spectrophotometric and electrochemical characterization. I. *Biochim. Biophys. Acta* **1057**, 239–257 (1991).
50. Teixeira, M. *et al.* Redox intermediates of *Desulfovibrio gigas* [NiFe] hydrogenase generated under hydrogen. Mossbauer and EPR characterization of the metal centers. *J. Biol. Chem.* **264**, 16435–16450 (1989).

Acknowledgements

This work was supported by grants from the NIH. We are grateful to C. A. Wraight, G. T. Babcock, W. Junge, R. J. P. Williams and H. B. Gray for timely encouragement and insight. Several crystallographers, most notably H. Michel, M. Frey, D. Rees, E. Berry and R. Huber, gave us crystal coordinates before publication.

Correspondence and requests for materials should be addressed to P.L.D. (e-mail: dutton@mail.med.upenn.edu). Assistance in calculating ρ and tunnelling rates can be found at our World-Wide Web site (www.med.upenn.edu/biocbiop/local_pages/dutton_lab.html).

Debris streams in the solar neighbourhood as relicts from the formation of the Milky Way

Amina Helmi*, Simon D. M. White†, P. Tim de Zeeuw* & HongSheng Zhao*

* Leiden Observatory, PO Box 9513, 2300 RA Leiden, The Netherlands

† Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85740 Garching bei München, Germany

It is now generally believed that galaxies were built up through gravitational amplification of primordial fluctuations and the subsequent merging of smaller precursor structures. The stars of the structures that assembled to form the Milky Way should make up much or all of its bulge and halo, in which case one hopes to find 'fossil' evidence for those precursor structures in the present distribution of halo stars. Confirmation that this process is continuing came with the discovery of the Sagittarius dwarf galaxy¹, which is being disrupted by the Milky Way, but direct evidence that this process provided the bulk of the Milky Way's population of old stars has hitherto been lacking. Here we show that about ten per cent of the metal-poor stars in the halo of the Milky Way, outside the radius of the Sun's orbit, come from a single coherent structure that was disrupted during or soon after the Galaxy's formation. This object had a highly inclined orbit about the Milky Way at a maximum distance of ~ 16 kpc, and it probably resembled the Fornax and Sagittarius dwarf spheroidal galaxies.

Early studies treated the formation of the Milky Way's spheroid as an isolated collapse, argued to have been either rapid and 'monolithic'², or inhomogeneous and slow compared to the motions of typical halo stars³. A second dichotomy distinguished 'dissipationless' galaxy formation, in which stars formed before collapse⁴, from 'dissipative' models in which the collapsing material was mainly gaseous⁵. Aspects of these dichotomies remain as significant issues in current theories⁶, but they are typically rephrased as questions about whether small units equilibrate and form stars before they are incorporated into larger systems, and about whether they are completely disrupted after such incorporation. Stars from Galactic precursors should be visible today either as 'satellite' galaxies, if disruption was inefficient, or as part of the stellar halo and bulge, if it was complete.

Recent work⁷ examined the present-day distribution expected for the debris of a precursor which was disrupted during or soon after the formation of the Milky Way. Objects which could contribute substantially to the stellar halo near the Sun must have had relatively short orbital periods. Ten Gyr after disruption their stars should be spread evenly through a large volume, showing none of the trails characteristic of currently disrupting systems like Sagittarius⁸. In any relatively small region, such as the solar neighbourhood, the stars should be concentrated into a number of coherent 'streams' in velocity space, each showing an internal-velocity dispersion of only a few km s^{-1} . Objects initially similar to the Fornax or Sagittarius dwarf galaxies should give rise to a few streams in the vicinity of the Sun.

The high-quality proper motions provided by the Hipparcos satellite allow us to construct accurate three-dimensional velocity distributions for almost-complete samples of nearby halo stars. On the basis of two recent observational studies^{9,10}, we define a sample containing 97 metal-deficient ($[\text{Fe}/\text{H}] \leq -1.6$ dex) red giants and RR Lyrae stars within 1 kpc of the Sun and with the following properties: (1) Hipparcos proper motions are available for 88 of them¹¹; for the remaining stars there are ground-based

measurements¹²; in all cases accuracies of a few mas yr^{-1} are achieved. (2) Radial velocities have been measured from the ground, with accuracies of the order of 10 km s^{-1} . Metal abundances have been determined either spectroscopically or from suitable photometric calibrations^{13–15}. (3) Calibrations of absolute magnitude M_V against $[\text{Fe}/\text{H}]$, for red giants^{13–15} and RR Lyrae stars¹⁶, allow photometric parallaxes to be derived to an accuracy of roughly 20%. These are more accurate than the corresponding Hipparcos trigonometric parallaxes, but still remain the largest source of uncertainty in the derived tangential velocities. (4) We estimate the completeness to be of the order of $\sim 92\%$, based on the fact that there are eight known giants which satisfy our selection criteria but do not have measured proper motions.

We look for substructure in our set of halo stars by studying the entropy S of the sample, defined as:

$$S = - \sum_i \frac{N_i}{N} \log \frac{N_i A_p}{N} \quad (1)$$

where the sum is over the A_p elements of the partition, the i th element contains N_i stars, and N is the total number of stars. In the presence of substructure the measured entropy will be smaller than that of a smooth distribution, and will depend on the details of the partition; some partitions will enhance the signal, whereas others will smear it out. If there is no substructure then all partitions will yield similar S values, and no significant minimum value will be found.

We implement this entropy test initially by partitioning velocity space into cubic cells 70 km s^{-1} on a side. This choice is a compromise. It leaves a large number of cells in the high-velocity range empty, but in the regions containing most of the stars, there are at least a few stars per cell.

It is necessary to quantify the significance of any observed low entropy value relative to the distribution expected in the absence of substructure. Here we do this by generating Monte Carlo realizations which test whether the kinematics of the sample are consistent with a multivariate gaussian distribution¹⁷. We calculate entropies for 10,000 Monte Carlo samples on the same partition as the real

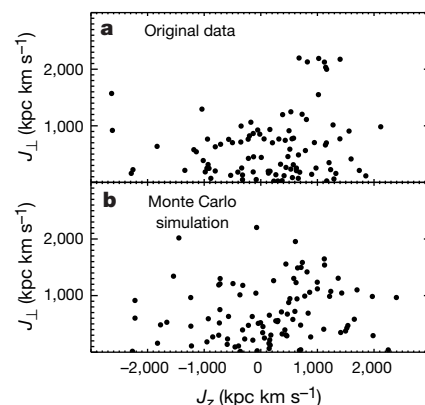


Figure 1 The distribution of nearby halo stars in the plane of angular momentum components, J_z versus $J_\perp = \sqrt{J_x^2 + J_y^2}$. Data are shown for our near-complete sample (a) and for one Monte Carlo realization (b). Our Monte Carlo data sets have the same number of stars and the same spatial distribution as the observed sample. The characteristics parameters of the multivariate gaussian used to describe the kinematics are obtained by fitting to the observed mean values and variances after appropriate convolution with the observational errors. We then generate 10,000 'observed' samples as follows. A velocity is drawn from the underlying multivariate gaussian; it is transformed to a proper motion and radial velocity (assuming the observed parallax and position on the sky); observational 'errors' are added to the parallax, the radial velocity and the proper motion; these 'observed' quantities are then transformed back to an 'observed' velocity. Velocities are referred to the Galactic Centre; we adopt 8 kpc as the distance to the Galactic Centre and 220 km s^{-1} towards galactic longitude $l = 0$ and galactic latitude $b = 0$ as the velocity of the local standard of rest.

data; only for 5.6% do we find values of S smaller than observed. We have repeated this test for many partitions, finding a large number with probabilities as low or lower than this. In particular, for a partition with a 250 km s^{-1} bin in v_ϕ and 25 km s^{-1} bins in v_R and v_z (v_R , v_ϕ and v_z are the velocity components in the radial, azimuthal and z -directions respectively), only 0.06% of Monte Carlo simulations have S smaller than observed. In general, cubic cells yield lower significance levels, suggesting that the detected structure may be elongated along v_ϕ . We conclude that a multivariate gaussian does not properly describe the distribution of halo star velocities in the solar neighbourhood.

At this point the main problem is to identify the structure which makes the observed data incompatible with a smooth velocity distribution. A comparison of the three principal projections of the observed distribution to similar plots for our Monte Carlo samples reveals no obvious differences. To better identify streams we turn to the space of adiabatic invariants. Here clumping should be stronger, as all stars originating from the same progenitor have very similar integrals of motion, resulting in a superposition of the corresponding streams. We focus on the plane defined by two components of the angular momentum: J_z and $J_\perp = \sqrt{J_x^2 + J_y^2}$, although $J_\perp$ is not fully conserved in an axisymmetric potential. In Fig. 1a we plot J_z versus $J_\perp$ for our sample. For comparison, Fig. 1b gives a similar plot for one of our Monte Carlo samples. For $J_\perp \leq 1,000 \text{ kpc km s}^{-1}$ and $|J_z| \leq 1,000 \text{ kpc km s}^{-1}$, the observed distribution appears fairly smooth. In this region we find stars with relatively low angular momentum and at all inclinations. In contrast, for $J_\perp \geq 1,000 \text{ kpc km s}^{-1}$, there are a few stars moving on retrograde low-inclination orbits, an absence of stars on polar orbits, and an apparent 'clump' on a prograde high-inclination orbit.

To determine the significance of this clumping, and to confirm it as the source of the signal detected by our entropy test, we compare the observed star counts in this plane to those for our Monte Carlo data sets. We count how many stars fall in each cell of a given partition of this angular momentum plane and compare it to the expected number in the Monte Carlo simulations. We say that the i th cell has a significant overdensity if there is less than 1% probability of obtaining a count as large as the observed N_i from a Poisson distribution with mean $\langle N^i \rangle = N_{\text{sim}}^{-1} \sum_{j=1}^{N_{\text{sim}}} N_j^i$, where N_j^i is the count in the i th cell of the j th simulation, and N_{sim} is the number of simulations. We repeat this test for a series of regular partitions of $p \times q$ elements, with p, q ranging from 3 to 20, thus allowing a clear

identification of the deviant regions. We find a very significant deviation in most partitions for cells with $J_\perp \approx 2,000 \text{ kpc km s}^{-1}$ and $500 < J_z < 1,500 \text{ kpc km s}^{-1}$; the probabilities of the observed occupation numbers range from 0.03% to 0.98%, depending on the partition, and in some partitions more than one cell is significantly overdense.

Given this apparently significant evidence for substructure in the local halo, we investigate what happens if we relax our metallicity and distance selection criteria. We proceed by including in our sample all red giants and RR Lyrae stars studied by Chiba and Yoshii¹⁰ with metallicities less than -1 dex and distances to the Sun of less than 2.5 kpc . This new sample contains 275 giant stars and adds 5 new stars to the most significant clump in our complete sample. Of the 13 members of the clump, 9 have $[\text{Fe}/\text{H}] \leq -1.6$, whereas the remaining 4 have $[\text{Fe}/\text{H}] \approx -1.53 \pm 0.12$, indicating that they are also very metal-poor. These stars are distributed all over the sky with no obvious spatial structure.

In Fig. 2 we highlight the kinematic structure of the clump in the extended sample. The clump stars are distributed in two streams moving in opposite directions perpendicular to the Galactic plane, with one possible outlier. This star has $v_R = 285 \pm 21 \text{ km s}^{-1}$, and we exclude it because its energy is too large to be consistent with the energies of the other members of the clump. The velocity dispersions for the stream with negative v_z (9 stars) are $\sigma_\phi = 30 \pm 17$, $\sigma_R = 105 \pm 16$ and $\sigma_z = 24 \pm 28 \text{ km s}^{-1}$, whereas for the stream with positive v_z (3 stars) these are $\sigma_\phi = 49 \pm 22$, $\sigma_R = 13 \pm 33$ and $\sigma_z = 31 \pm 28 \text{ km s}^{-1}$. An elongation in the v_R direction is expected for streams close to their orbital pericentre (the closest distance to the Galactic Centre; compare with other plots of simulated streams⁷).

The orbit of the progenitor system is constrained by the observed positions and velocities of the stars. The orbital radii at apocentre and pericentre are $r_{\text{apo}} \approx 16 \text{ kpc}$ and $r_{\text{peri}} \approx 7 \text{ kpc}$, the maximum height above the plane is $z_{\text{max}} \approx 13 \text{ kpc}$, and the radial period is $P \approx 0.4 \text{ Gyr}$, for a Galactic potential including a disk, bulge and dark halo⁸. We run numerical simulations of satellite disruption in this potential to estimate the initial properties of the progenitor. After 10 Gyr of evolution, we find that the observed properties of the streams detected can be matched by stellar systems similar to dwarf spheroidals with initial velocity dispersions σ in the range $12\text{--}18 \text{ km s}^{-1}$ and core radii R of $0.5\text{--}0.65 \text{ kpc}$. We also analysed whether the inclusion of an extended dark halo around the initial object would affect the structures observed and found very little effect. We

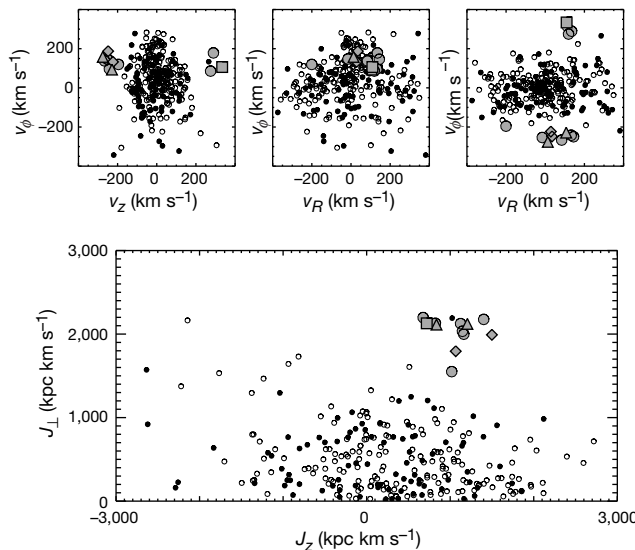


Figure 2 The distribution of nearby halo stars in velocity space and in the J_z – $J_\perp$ plane. Data are shown for our original sample (filled circles) and for the extended sample of more metal-rich and more distant giants¹⁰ (open circles). Candidates for our detected

substructure are highlighted in grey: triangles indicate more metal-rich giant stars at distances $>1 \text{ kpc}$, diamonds indicate more metal-rich giants at $\leq 1 \text{ kpc}$, squares indicate metal-poor giants at $>1 \text{ kpc}$, and circles indicate metal-poor giants at $\leq 1 \text{ kpc}$.

derive the initial luminosity L from $L = L^*/(f^{\text{giant}} C^* f^{\text{sim}})$, where $L^* = 350 L_{\odot}$ is the total luminosity of the giants in the clump in our near-complete sample, $f^{\text{giant}} \approx 0.13$ is the ratio of the luminosity in giants with M_V and $(B - V)$ in the range observed to the total luminosity of the system for an old metal-poor stellar population¹⁸, $C^* = 0.92$ is our estimated completeness, and $f^{\text{sim}} \approx 1.9 \times 10^{-4}$ is the fraction of the initial satellite contained in a sphere of 1 kpc radius around the Sun as determined from our simulations. This gives $L \approx 1.5 \times 10^7 L_{\odot}$, from which we can derive, using our previous estimates of the initial velocity dispersion and core radii, an average initial core mass-to-light ratio $M/L \approx 3\text{--}10 \Upsilon_{\odot}$, where $\Upsilon_{\odot}$ is the mass-to-light ratio of the Sun. A progenitor system with these characteristics would be similar to Fornax. Moreover, the mean metal abundance of the stars is consistent with the derived luminosity, if the progenitor follows the known metallicity–luminosity relation of dwarf satellites in the Local Group¹⁹.

The precursor object was apparently on an eccentric orbit with a relatively large apocentre. Given that it contributes 7/97 of the local halo population, our simulations suggest that it should account for 12% of all metal-poor halo stars outside the solar circle. Figure 2 shows that there are few other halo stars on high angular-momentum polar orbits in the solar neighbourhood, just the opposite of the observed kinematics of satellites of the Milky Way²⁰. The absence of satellite galaxies on eccentric non-polar orbits argues that some dynamical process preferentially destroys such systems; their stars should then end up populating the stellar halo. As we have shown, the halo does indeed contain fossil streams with properties consistent with such disruption. $\square$

Received 18 June; accepted 13 September 1999.

1. Ibata, R., Gilmore, G. & Irwin, M. J. A dwarf satellite galaxy in Sagittarius. *Nature* **370**, 194–196 (1994).
2. Eggen, O. J., Lynden-Bell, D. & Sandage, A. R. Evidence from the motions of old stars that the Galaxy collapsed. *Astrophys. J.* **136**, 748–766 (1962).
3. Searle, L. & Zinn, R. Compositions of halo clusters and the formation of the galactic halo. *Astrophys. J.* **225**, 357–379 (1978).
4. Gott, J. R. Recent theories of galaxy formation. *Annu. Rev. Astron. Astrophys.* **15**, 235–266 (1977).
5. Larson, R. B. Models for the formation of elliptical galaxies. *Mon. Not. R. Astron. Soc.* **173**, 671–699 (1975).
6. White, S. D. M. & Frenk, C. S. Galaxy formation through hierarchical clustering. *Astrophys. J.* **379**, 52–79 (1991).
7. Helmi, A. & White, S. D. M. Building up the stellar halo of the Galaxy. *Mon. Not. R. Astron. Soc.* **307**, 495–517 (1999).
8. Johnston, K. V., Hernquist, L. & Bolte, M. Fossil signatures of ancient accretion events in the Halo. *Astrophys. J.* **465**, 278–287 (1996).
9. Beers, T. C. & Sommer-Larsen, J. Kinematics of metal-poor stars in the Galaxy. *Astrophys. J. Suppl. Ser.* **96**, 175–221 (1995).
10. Chiba, M. & Yoshii, Y. Early evolution of the Galactic halo revealed from Hipparcos observations of metal-poor stars. *Astron. J.* **115**, 168–192 (1998).
11. *The Hipparcos and Tycho Catalogues* (SP-1200, European Space Agency, ESA Publications Division, ESTEC, Noordwijk, The Netherlands, 1997).
12. Roeser, S. & Bastian, U. A new star catalogue of SAO type. *Astron. Astrophys. (suppl.)* **74**, 449–451 (1988).
13. Anthony-Twarog, B. J. & Twarog, B. A. Reddening estimation for halo red giants using uvby photometry. *Astron. J.* **107**, 1577–1590 (1994).
14. Beers, T. C., Preston, G. W., Shectman, S. A. & Kage, J. A. Estimation of stellar metal abundance. I—Calibration of the Ca II K index. *Astron. J.* **100**, 849–883 (1990).
15. Norris, J., Bessell, M. S. & Pickles, A. J. Population studies. I. The Bidelman-MacConnell “weak-metal” stars. *Astrophys. J. Suppl. Ser.* **58**, 463–492 (1985).
16. Layden, A. C. The metallicities and kinematics of RR Lyrae variables I: New observations of local stars. *Astron. J.* **108**, 1016–1041 (1994).
17. Sommer-Larsen, J., Beers, T. C., Flynn, C., Wilhelm, R. & Christensen, P. R. A dynamical and kinematical model of the Galactic stellar halo and possible implications for Galaxy formation scenarios. *Astrophys. J.* **481**, 775–781 (1997).
18. Bergbusch, P. A. & Vandenberg, D. A. Oxygen-enhanced models for globular cluster stars. II. Isochrones and luminosity functions. *Astrophys. J. Suppl. Ser.* **18**, 163–220 (1992).
19. Mateo, M. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).
20. Lynden-Bell, D. & Lynden-Bell, R. M. Ghostly streams from the formation of the Galaxy’s halo. *Mon. Not. R. Astron. Soc.* **275**, 429–442 (1995).

Acknowledgements

A.H. thanks the Max-Planck-Institut für Astrophysik for hospitality during her visits. We made use of the Simbad database (maintained by Centre de Données astronomiques de Strasbourg) and of the HIPPARCOS online facility at the European Space Research and Technology Centre (ESTEC) of the European Space Agency (ESA).

Correspondence and requests for materials should be addressed to A.H. (e-mail: ahelmi@strw.leidenuniv.nl).

Multiple stellar populations in the globular cluster ω Centauri as tracers of a merger event

Y.-W. Lee*, J.-M. Joo*, Y.-J. Sohn*, S.-C. Rey*, H.-c. Lee* & A. R. Walker†

* Center for Space Astrophysics, Yonsei University, Seoul 120-749, Korea

† National Optical Astronomy Observatories/Cerro Tololo Interamerican Observatory (NOAO/CTIO), Casilla 603, La Serena, Chile

The discovery of the Sagittarius dwarf galaxy¹, which is being tidally disrupted by and merging with the Milky Way, supports the view that the halo of the Galaxy has been built up at least partially by the accretion of similar dwarf systems. The Sagittarius dwarf contains several distinct populations of stars^{2,3}, and includes M54 as its nucleus, which is the second most massive globular cluster associated with the Milky Way. The most massive globular cluster is ω Centauri, and here we report that ω Centauri also has several distinct stellar populations, as traced by red-giant-branch stars. The most metal-rich red-giant-branch stars are about 2 Gyr younger than the dominant metal-poor component, indicating that ω Centauri was enriched over this timescale. The presence of more than one epoch of star formation in a globular cluster is quite surprising, and suggests that ω Centauri was once part of a more massive system that merged with the Milky Way, as the Sagittarius dwarf galaxy is in the process of doing now. Mergers probably were much more frequent in the early history of the Galaxy and ω Centauri appears to be a relict of this era.

As part of our investigation of the luminosity–metallicity relation of the RR Lyrae stars in the globular cluster ω Cen, we have obtained 2K BV CCD (charge-coupled device) frames with the CTIO 0.9-m

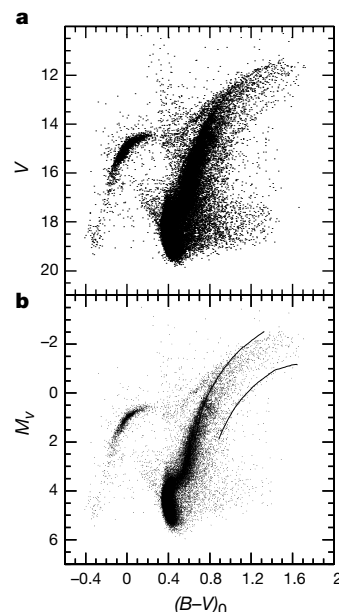


Figure 1 Colour–magnitude diagram of 50,129 stars in the direction of ω Cen. These diagrams were obtained from a mosaic of nine 2K CCD (charge-coupled device) fields. Only stars with at least 20 detections and small photometric errors ($\sigma_V < 0.05$ mag and $\sigma_{B-V} < 0.071$ mag) have been plotted. **a**, All stars in our programme field; **b**, stars located between 2.58 and 15.48 arcmin from the cluster centre. There are several distinct RGBs and a red clump associated with the most metal-rich component. Two RGB loci from the new Yale isochrones⁶ (metallicity $Z = 0.0004, 0.005$) are also compared (**b**, solid lines) that bracket the metallicity range of ω Cen. M_V , absolute visual magnitude.

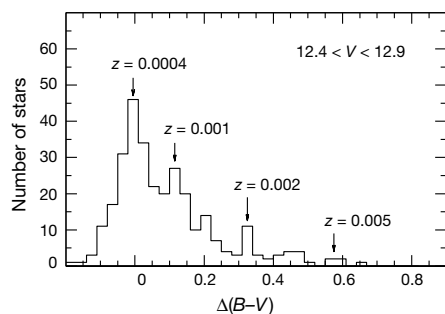


Figure 2 Histogram of the distribution of colour difference. Colour difference between each RGB star and the RGB fiducial of the most metal-poor component, $\Delta(B-V)$, is plotted in the range $12.4 < V < 12.9$. The metallicities (Z values) of four distinct RGBs are also marked.

telescope that cover 40×40 arcmin² in a 3×3 grid centred on the cluster, and covering out to approximately half the tidal radius. In total, 40–42 frames were taken in each filter and each field. The seeing was between 1.0 and 1.7 arcsec, and all of the observing nights (5–10 April 1996) were fully photometric. The B and V magnitudes of individual stars were measured with the point-spread-function (PSF)-fitting programs DAOPHOT II and ALLSTAR in the standard manner⁴. As a by-product of this investigation, we obtained high-quality homogeneous BV colour–magnitude data for more than 130,000 stars in the field toward ω Cen, which represents the most extensive photometric survey to date for this cluster.

Figure 1 shows a V versus $B-V$ colour–magnitude diagram for stars in our programme field. We note the presence of several distinct red-giant branches (RGBs) with a red, presumably metal-rich, sequence well separated from other bluer metal-poor ones. This feature was not evident in previous photometry⁵ with smaller sample sizes and larger photometric uncertainties. The radial distribution of the most metal-rich RGB stars is not significantly different from those of the metal-poor ones, which confirms that they belong to ω Cen. Note also that the red clump that must be associated with the most metal-rich RGB is clearly apparent, partially overlapping the metal-poor RGBs. The presence of other interesting features on the colour–magnitude diagram, such as the blue-tail phenomenon of the horizontal-branch and the blue straggler stars, illustrates the diversity of stellar populations in this cluster. The signature of field-star contamination is also evident, primarily as a swathe of stars with $0.3 \leq (B-V)_0 \leq 1.2$. These stars will belong to the foreground galactic disk population.

In order to further investigate the discrete nature of the RGB, we have plotted in Fig. 2 a histogram of the distribution of colour difference between each RGB star and the RGB fiducial of the most metal-poor component. The RGB stars are selected in a relatively narrow magnitude range $12.4 < V < 12.9$, so that the field-star and red-clump contamination is minimized, and which also avoids artificial mixing in the histogram stemming from metallicity dependence of the RGB slope. The presence of several distinct RGBs is confirmed, although the most metal-rich component is not as well distinguished as in Fig. 1 because of the small sample size of such stars in this magnitude range.

Figure 3 shows our population models. These illustrate the relative age estimation from the location of the red clump associated with the most metal-rich RGB with respect to blue horizontal-branch stars associated with the most metal-poor component. Figure 3a shows the case where all stars have the same age despite their different metallicities, which produces a red horizontal branch that is clearly much bluer than the observed red clump confined within the two most extreme RGBs. Figure 3b shows our best match with the observed colour–magnitude diagram in Fig. 1, which

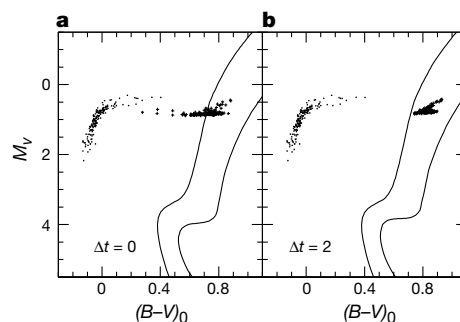


Figure 3 Stellar population models. The models illustrate the estimation of age difference between the red clump associated with the most metal-rich ($Z = 0.005$) RGB and the blue horizontal-branch associated with the most metal-poor ($Z = 0.0004$) component. Dots and crosses are individual horizontal-branch stars from synthetic horizontal-branch models⁷ and the solid lines are from new Yale isochrones⁶. **a**, All stars have the same age despite their different metallicities; **b**, the most metal-rich population is 2 Gyr younger than the most metal-poor population. Only a Δt of about 2 Gyr reproduces the features on the observed colour–magnitude diagram in Fig. 1.

suggests that the most metal-rich population in ω Cen is some 2 Gyr younger than the most metal-poor population in this system. This internal age–metallicity relation is clear evidence that ω Cen has enriched itself over this timescale.

The multimodal metallicity distribution function and the age–metallicity relation found in this study suggest that the protocluster of ω Cen was massive enough to undergo some self-enrichment and several early bursts of star formation. The relatively extended enrichment period of about 2 Gyr then indicates that the initial evolution of ω Cen occurred away from the dense central regions of the young Galaxy, for if this were not the case, one would expect the gaseous material to have been stripped from the cluster on a much shorter timescale⁸. The plausible resolution of this problem is that ω Cen has evolved within a dwarf-galaxy-sized gas-rich subsystem (the Searle and Zinn⁹ fragment) until it merged with and was disrupted by our Galaxy some 2 Gyr after the formation of its first-generation metal-poor stars, leaving its core as today's globular cluster ω Cen. This is consistent with the fact that the multiple populations observed in ω Cen bear a strong resemblance to the Sagittarius dwarf system, where three distinct RGBs are identified with an internal age–metallicity relation that spans more than 3 Gyr (refs 2, 3). Alternatively, it might be argued that a merger of several clusters could also create a composite object with multiple populations. The well defined peaks in the metallicity distribution function in Fig. 2, however, require a merger of at least four clusters, and the possibility of this occurring may be extremely low. Furthermore, cluster mergers could only have occurred in dwarf galaxies with velocity dispersions much smaller than the Milky Way's¹⁰. Thus, even if ω Cen was formed by cluster mergers, it is probably still a relict of a dwarf galaxy subsequently accreted by our Galaxy. If our interpretation is correct, the case of ω Cen and that of the Sagittarius dwarf system would provide direct evidence for past and continuing accretion of protogalactic fragments, which suggest that similar accretion events may have continued throughout the history of Galactic formation. Extensive photometric surveys for other massive globular clusters, especially those with peculiar colour–magnitude diagram morphology, such as M22 with a broad RGB like ω Cen, and NGC2808 with bimodal horizontal-branch distribution, will undoubtedly help to reveal whether they also represent the relicts of the Galaxy building blocks. □

Received 29 June; accepted 14 September 1999.

1. Ibata, R. A., Gilmore, G. & Irwin, M. J. A dwarf satellite galaxy in Sagittarius. *Nature* **370**, 194–196 (1994).

2. Sarajedini, A. & Layden, A. C. A photometric study of the globular cluster M54 and the Sagittarius dwarf galaxy: Evidence for three distinct populations. *Astron. J.* **109**, 1086–1094 (1995).
3. Layden, A. C. & Sarajedini, A. C. The globular cluster M54 and the star formation history of the Sagittarius dwarf galaxy. *Astrophys. J.* **486**, L107–L110 (1997).
4. Stetson, P. B. DAOPHOT: A computer program for crowded-field stellar photometry. *Publ. Astron. Soc. Pacif.* **99**, 191–222 (1987).
5. Woolley, R. v. d. Studies of the globular cluster ω Centauri. I. Catalogue of magnitudes and proper motions. *R. Observ. Ann.* No. 2, 1–128 (1966).
6. Demarque, P. et al. *New Yale Isochrones* (Yale University Observatory, New Haven, 1996).
7. Lee, Y.-W., Demarque, P. & Zinn, R. The horizontal-branch stars in globular clusters. II. The second parameter phenomenon. *Astrophys. J.* **423**, 248–265 (1994).
8. Norris, J. E., Freeman, K. C. & Mighell, K. J. The giant branch of ω Centauri. V. The calcium abundance distribution. *Astrophys. J.* **462**, 241–254 (1996).
9. Searle, L. & Zinn, R. Compositions of halo clusters and the formation of the galactic halo. *Astrophys. J.* **225**, 357–379 (1978).
10. van den Bergh, S. Mergers of globular clusters. *Astrophys. J.* **471**, L31–L32 (1996).

Acknowledgements

Support for this work was provided by the Creative Research Initiatives Program of the Korean Ministry of Science & Technology, and in part, by the Korea Science & Engineering Foundation. S.-C.R. was a visiting astronomer at CTIO/NOAO, which is operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

Correspondence and requests for materials should be addressed to Y.-W.L. (e-mail: ywlee@csa.yonsei.ac.kr).

Discovery of a planet orbiting a binary star system from gravitational microlensing

D. P. Bennett*, S. H. Rhie*, A. C. Becker†, N. Butler*, J. Dann‡, S. Kaspi‡, E. M. Leibowitz‡, Y. Lipkin‡, D. Maoz‡, H. Mendelson‡, B. A. Peterson‡, J. Quinn*, O. Shemmer‡, S. Thomson|| & S. E. Turner¶

* Physics Department, University of Notre Dame, Notre Dame, Indiana 46530, USA

† Departments of Physics and Astronomy, University of Washington, Seattle, Washington 98195, USA

‡ School of Physics & Astronomy and Wise Observatory, Tel-Aviv University, Tel-Aviv 69978, Israel

§ Mt Stromlo and Siding Spring Observatory, Australian National University, Weston, Australian Capital Territory, Australia

|| Department of Mathematics & Statistics, Monash University, Clayton, Victoria 3168, Australia

¶ Department of Physics, University of California, Los Angeles, California 90095, USA

The properties of the recently discovered^{1,2} extrasolar planets were not anticipated by theoretical work on the formation of planetary systems, most models for which were developed to explain our Solar System. Indeed, the observational technique used to detect these planets (measurement of radial-velocity shifts in stellar spectral lines) do not yet have the sensitivity to detect planetary systems like our own³. Here we report observations and modelling of the gravitational microlensing event MACHO-97-BLG-41. We infer that the lens system consists of a planet of about 3 Jupiter masses orbiting a binary stellar system consisting of a late-K dwarf star and an M dwarf. The stars are separated by ~ 1.8 astronomical units (1 AU is the Earth–Sun distance), and the planet is orbiting them at a distance of about 7 AU. We had expected to find first the microlensing signature of jovian planets around single stars, so this result suggests that such planets orbiting short-period binary stars may be common.

The Microlensing Planet Search (MPS) Collaboration aims to detect planets that orbit distant stars by detecting the influence of such planets on the light curves of gravitational microlensing events.

Gravitational microlensing events are observed via the time-varying magnification of the sum of the gravitationally lensed images as the lens system passes in front of the background source star. The separation of the images is too small to observe with current instruments. A planet orbiting the lens star can be detected by a brief deviation of the microlensing light curve⁴ from the normal single lens light curve^{5–7}.

The microlensing event MACHO-97-BLG-41 was discovered by the MACHO team⁸ and announced on 19 June 1997. MPS observations from the Mt Stromlo 1.9-m telescope began on that night. On 29 June, the MACHO team issued a further announcement that the light curve of this event did not have the shape expected for a single lens event, and the PLANET team issued a similar announcement on 2 July. Regular observations by the GMAN follow-up team began shortly after this second MACHO announcement, with nightly observations from the CTIO 0.9-m telescope and less frequent observations from the Wise Observatory 1.0-m telescope. The MACHO and GMAN data were previously presented without analysis by the MACHO/GMAN group⁸. Here we present a combined analysis of the MPS and MACHO/GMAN data.

The modelling of MACHO-97-BLG-41 has proven to be more difficult than any other multiple lensing event observed by MACHO/GMAN or MPS to date. Figure 1 shows the light curve of the combined data set along with the best-fit model, while Fig. 2

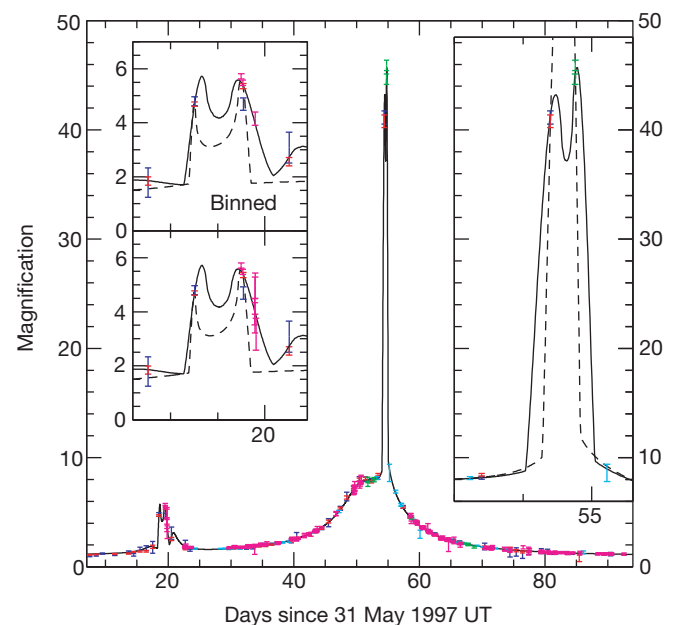


Figure 1 The light curve and best-fit model for the MACHO-97-BLG-41 microlensing event as a function of time. The data consist of 356 MPS R-band observations from the Mt Stromlo 1.9-m telescope, 197 MACHO-R and 194 MACHO-V band observations from the Mt Stromlo 1.3-m telescope, 35 R-band observations from the CTIO 0.9-m telescope, and 17 R-band observations from the Wise 1.0-m telescope. The MACHO-R, MACHO-V, Wise-R, CTIO-R and MPS data are plotted in red, blue, green, cyan and magenta, respectively. Insets show close-ups of the caustic crossing regions of the light curves; the tick interval for these figures is 1 day. The last 5 observations from the night of 20 June were short exposures taken in bright moonlight over a span of 40 min. These have been averaged into a single data point for the inset at upper left. The MPS and MACHO data were reduced with the SoDophot photometry routine²², while the CTIO and Wise data were reduced with ALLFRAME²³. The MACHO, CTIP and Wise data were previously presented by Alcock et al.⁸. The solid curve is the best-fit triple lens model described in the text, and the dashed curve in the inset figures is the 'best fit' orbiting binary lens light curve which requires an 'orbital velocity' larger than the escape velocity of the binary system and is therefore unphysical. This large orbital velocity is due to the large distance that the binary stars must move to account for both caustic crossing features and the small t_* value for this fit which is required to achieve the magnifications observed on 20–21 June.

shows the trajectory of the source star with respect to the lens masses, and the overall lens system caustic structure. Most of the features of these multiple lens light curves can be understood by inspection of diagrams like Fig. 2, because the regions of high magnification are associated with the caustic curves. The spikes and U-shaped light curves that commonly occur in binary lens lightcurves⁸ are easily understood as caustic crossing features.

Our initial attempts to fit the MACHO-97-BLG-41 light curve focused on both static and orbiting binary lens models. Static binary lens models have been successful in fitting the light curves of all the other known multiple lens events^{8–11}, but such a model is unable to account for the two caustic crossing features of MACHO-97-BLG-41. A static binary lens model would necessarily have a light curve peak around 21 July that is much larger than what is actually observed, because a cusp of the central caustic must point in the direction of the caustic that was crossed on 20–21 June. The inclusion of binary lens possible orbital motion¹² also failed to generate an acceptable fit.

When we concluded that binary lens models would not fit the data, we turned to triple lens models (S.H.R. and D.P.B., manuscript in preparation), and our fitting code quickly converged to a model which appears to correspond to a stable dynamical system. The light curve for this model is shown in Fig. 1. The model parameters are expressed in terms of the Einstein ring radius, R_E , which is the size of the ring image that would be seen with perfect lens–source alignment (that is, with the source and all the lens masses in a line). For a typical Galactic bulge lens system, $R_E \approx 3$ AU. The fit parameters are as follows: the Einstein ring radius crossing time is $t_E = 27.832$ days. The closest approach between the angular positions of the source and the lens system centre of mass is $u_{\min} = 0.0679$ (in units of R_E) which occurs at time $t_0 = 23.593$ July 1999, UT. The mass fractions

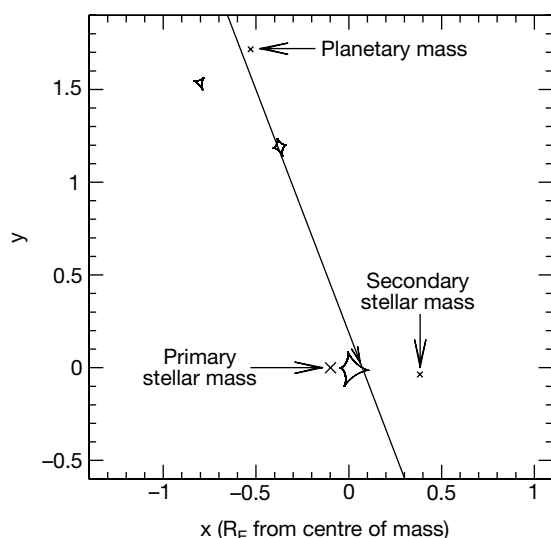


Figure 2 Caustic curves for the best-fit model for the MACHO-97-BLG-41 microlensing event, the lens positions, and the source trajectory. The last is the diagonal line running from top left to bottom right. High magnification occurs when the source passes inside a caustic curve where the magnification is proportional to the inverse square root of the distance to the caustic from the interior. High magnification also occurs outside a caustic curve in the vicinity of a caustic curve cusp. The observed light curve of MACHO-97-BLG-41 has the following features that can be seen in this figure and in Fig. 1. (1) The passage of an isolated caustic curve on 19–21 June, (2) an approach to a cusp on 21 July that gives rise to the light curve shoulder, (3) an extremely bright spike indicating the crossing of a very narrow caustic structure during 24–25 July. The triangular caustic curve located near the planetary mass is due to the two stellar mass lenses and has no effect on the light curve, although a similar caustic does play a role in the unphysical orbiting binary lens fit. The orbiting binary lens fit allows this triangular caustic to intersect the source trajectory to produce the 19–20 June caustic crossings, but only if the orbital velocity exceeds the escape velocity of the binary system.

of the three masses are 0.7870, 0.2086 and 0.0044, and the separations are $0.4822 R_E$ and $1.768 R_E$ between the mass pairs 1–2 and 1–3, respectively. The opening angle between the mass 1–2 and mass 1–3 vectors is 108.30° , and the source trajectory intersects the lens 1–2 axis at an angle of -64.81° . The measured widths and magnifications at the caustic crossings allow us to determine the time for the source star to move by its own radius with respect to the lens axis: $t_* = 0.16$ days, assuming linear limb darkening parameters of 0.656 for the standard R band, 0.632 for the MACHO-R band, and 0.777 for the MACHO-V band¹³ based on a spectrum of the 97-BLG-41 source star¹⁴. The relative locations of the lens masses, the caustics, and the source trajectory are shown in Fig. 2. This model also indicates that 10–20% of the flux (depending on the pass band) identified with the source star is unlensed, indicating that one or more unlensed stars lie within 1–2" of the lensed source. Such blending is quite common in such crowded stellar fields. Our model has a $\chi^2 = 1,102.96$ for 779 degrees of freedom, which is typical of other microlensing events without binary or planetary light curve deviations. The best-fit orbiting binary source model has a χ^2 that is larger by 67.57, and an implied orbital velocity larger than the maximum escape velocity of the binary system.

Orbital stability is an important concern for triple star systems, because many such systems are dynamically unstable. Our gravitational microlensing model indicates that the planet is separated from the stellar system centre of mass by 3.9 times the binary star separation, in the plane perpendicular to the line of sight. A recent stability study by Holman and Wiegert¹⁵ indicates that the stability of planets orbiting binary star systems depends on the stellar system orbital eccentricity. For our best-fit mass ratio and assuming a circular stellar orbit, they find that planetary orbits are stable if their median separation is at least 2.2 times that of the stars; for an

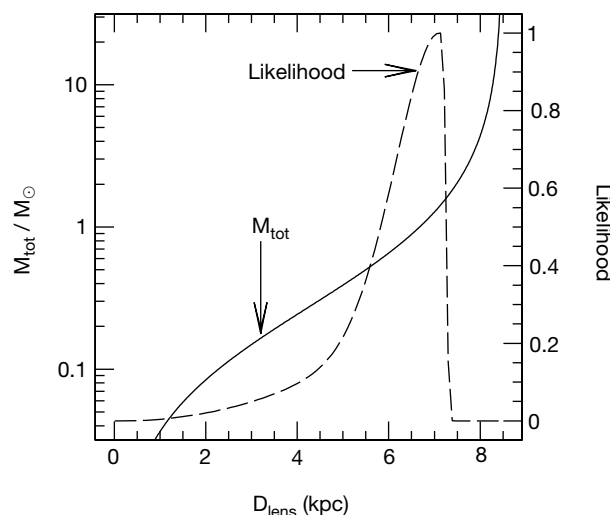


Figure 3 The mass–distance relation for the MACHO-97-BLG-41 lens system, and a likelihood function for the lens-system distance. The measurements of μ and t_E (see text) provide two constraints on the three unknowns of a microlensing event (the lens distance, mass, and transverse velocity), which allows us to solve for the single-parameter family of solutions shown here. The transverse velocity of the lens is also determined as a function of the distance to the lens. Taking the velocity distributions from a standard Galactic model²⁴, we are able to construct a likelihood function for the distance and mass of the lens system. For a lens distance ≥ 7 kpc, the primary lens must be more massive than the Sun and will contribute a significant amount of unlensed light superimposed on the source star. Using an upper limit on the brightness of the primary lens star of $20 \pm 10\%$ of the source brightness and assuming that the V-band brightness of a main-sequence star is proportional to $M^{4.5}$ (which is appropriate for $M \geq 1 M_\odot$), we find the likelihood function shown here. The very steep cut-off at $D_{\text{lens}} > 7$ kpc is due to this constraint on the brightness of the lens. Our main-sequence assumption could be avoided if the primary lens was a white dwarf, but as white dwarfs with masses $> 1 M_\odot$ are rare, it is unlikely that our limits on the mass and location of the lens system are violated.

eccentricity as high as 0.7, the critical ratio of the median separations climbs to about 3.1. Thus, our triple lens fit appears to correspond to a stable system as long as the separation ratio is not substantially reduced when the unobserved line-of-sight component is added.

Further details about the properties of the lens system can be obtained by comparing the source star radius crossing time, t_* , to the properties of this star as determined by Lennon *et al.*¹⁴. Their spectrum indicates an effective temperature $T_{\text{eff}} = 5,000 \pm 200$ K, a surface gravity $\log g = 3.2 \pm 0.3$, and an iron to hydrogen ratio $[\text{Fe}/\text{H}] = -0.2 \pm 0.3$ dex, while our model and the MACHO Project's photometric calibration¹⁶ indicate that the unlensed magnitude and colour of the source star are $V = 19.66 \pm 0.20$ and $V - R = 1.24 \pm 0.10$. This allows us to estimate the extinction to be $A_V = 2.9 \pm 0.4$, which leads to an estimate of the angular size of the source star: $\theta_* = 2.9 \pm 0.7 \mu\text{as}$. The distance to the source star is not known precisely, but it is likely to be at, or slightly beyond, the centre of the Galaxy at $D \approx 8.5$ kpc. At this distance, the source star would have a radius of $5.3 \pm 1.3 R_\odot$, where $R_\odot$ is the solar radius. A comparison to the Bertelli *et al.*¹⁷ isochrones indicates that a number of possible early K class III–IV stellar models are consistent with these parameters. These range from a 10-billion-year-old $1 M_\odot$ star to a 1.5-billion-year-old star of $1.7 M_\odot$ (here $M_\odot$ is the solar mass).

Our estimate of the angular size of the source star, combined with our measurement of the stellar radius crossing time, yields the relative proper motion of the lens with respect to the source: $\mu = 6.7 \pm 1.7 \text{ mas yr}^{-1}$. Projected to the likely source distance of 8.5 kpc, this is 270 km s^{-1} which is a reasonable value for the relative transverse velocity between a pair of Galactic-bulge stars. The measurements of μ and t_E allow us to solve for the single parameter family of solutions relating lens mass and distance shown in Fig. 3. This figure also shows the likelihood function for the lens system location, which is based on a simple model of Galactic kinematics and our upper limit on the lens brightness.

The likelihood function in Fig. 3 implies a lens distance $D_{\text{lens}} = 6.3^{+0.6}_{-1.3}$ kpc and a total mass $M_{\text{tot}} = 0.8 \pm 0.4 M_\odot$. The individual lens masses are $M_1 = 0.6 \pm 0.3 M_\odot$, $M_2 = 0.16 \pm 0.08 M_\odot$, and $M_3 = 0.0033 \pm 0.0017 M_\odot$ (or $M_3 = 3.5 \pm 1.8$ Jupiter masses). The implied transverse separations are $1.5^{+0.1}_{-0.3}$ AU for the two stars, and $5.7^{+0.6}_{-1.1}$ AU for the planet from the centre of mass of the system. Assuming a random orientation, the most likely three-dimensional separations are a stellar separation of 1.8 AU, and a planetary separation of 7.0 AU. Thus, the most likely lens system consists of a K-dwarf/M-dwarf binary star pair orbited by a planet of about 3 Jupiter masses.

One alternative explanation of this event might be a binary source star system lensed by a binary lens system¹⁸. This would require the secondary source star to be much fainter than the primary, and to be lensed by a much greater amount, so that the width of the secondary source light curve will be much narrower than the primary source light curve. In general, one would expect a significant colour shift between the secondary and primary peaks in such a situation. The MACHO data show no evidence for such a colour shift. The PLANET Collaboration apparently has additional data on this event¹⁹ in the I band and perhaps in the V-band as well, so future attempts to model the light curve may be able to place tighter constraints on any colour change. A definitive test of the binary source–binary lens hypothesis would be to search for radial velocity variability of the source star due to its orbital motion.

Our apparent detection of a planet orbiting a binary star system represents, to our knowledge, both the first discovery of a jovian planet via gravitational microlensing and the first discovery of a planet orbiting a binary star system. (The MPS and MOA²⁰ teams have recently reported a possible detection of a low-mass planet via microlensing. A number of planets have been discovered orbiting individual members of widely separated binary systems², but these

are systems that probably became binaries by gravitational capture after the planets had formed. The planet orbiting the MACHO-97-BLG-41 lens system is likely to be the first example of a planet that formed in a binary system, although circumbinary disks that might be in the process of forming planets have been observed²¹. It is curious that the planet we have detected via gravitational microlensing should apparently be orbiting a binary star system. Such planets have not been seen in planetary searches using the radial velocity technique, because the radial velocity search programmes have avoided short-period binary stars. Microlensing planet search programmes tend to concentrate on binary lensing events, because they provide better opportunities for non-planetary science^{8,9}. As these events comprise $\approx 10\%$ of the total number of microlensing events, many more single-lens events are observed, and it was expected that microlensing would first detect jovian planets orbiting single stars. Our present result suggests that jovian planets may be more common in short-period binary systems than in single star systems. $\square$

Received 5 August; accepted 4 October 1999.

- Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–357 (1995).
- Marcy, G. W. & Butler, R. P. Detection of extrasolar giant planets. *Annu. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
- Cumming, A., Marcy, G. W. & Butler, R. P. The Lick planet search: detectability and mass thresholds. *Astrophys. J.* (in the press); preprint astro-ph/9906466 at (<http://xxx.lanl.gov>) (1999).
- Mao, S. & Paczyński, B. Gravitational microlensing by double stars and planetary systems. *Astrophys. J.* **374**, L37–L41 (1991).
- Aubourg, E. *et al.* Evidence for gravitational microlensing by dark objects in the galactic halo. *Nature* **365**, 623–625 (1993).
- Alcock, C. *et al.* Possible gravitational microlensing of a star in the Large Magellanic Cloud. *Nature* **365**, 621–623 (1993).
- Udalski, A. *et al.* The optical gravitational lensing experiment. Discovery of the first candidate microlensing event in the direction of the galactic bulge. *Acta Astron.* **43**, 289–293 (1993).
- Alcock, C. *et al.* Binary microlensing events from the MACHO project. *Astrophys. J.* (submitted); preprint astro-ph/9907369 at (<http://xxx.lanl.gov>) (1999).
- Albrow, M. *et al.* Limb-darkening of a K giant in the galactic bulge: PLANET photometry of MACHO 97-BLG-28. *Astrophys. J.* (in the press); preprint astro-ph/9811479 at (<http://xxx.lanl.gov>) (1999).
- Udalski, A. *et al.* The optical gravitational lensing experiment: OGLE no. 7: binary microlens or a new unusual variable? *Astrophys. J.* **436**, L103–L106 (1994).
- Alard, C., Mao, S. & Guibert, J. Object DUO 2: a new binary lens candidate? *Astron. Astrophys.* **300**, L17–L20 (1995).
- Dominik, M. Galactic microlensing with rotating binaries. *Astron. Astrophys.* **329**, 361–374 (1998).
- Diaz-Cordoves, J., Claret, A. & Gimenez, A. Linear and non-linear limb-darkening coefficients for LTE model atmospheres. *Astron. Astrophys. Suppl. Ser.* **110**, 329–350 (1995).
- Lennon, D. J. *et al.* Real-time spectroscopy of gravitational microlensing events—probing the evolution of the galactic bulge. *ESO Messenger* **90**, 30–38 (1997); preprint astro-ph/9711147 at (<http://xxx.lanl.gov>) (1999).
- Holman, M. J. & Wiegert, P. A. Long-term stability of planets in binary systems. *Astron. J.* **117**, 621–628 (1999).
- Alcock, C. *et al.* Calibration of the MACHO photometry database. *Proc. Astron. Soc. Pac.* (in the press).
- Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F. & Nasi, E. Theoretical isochrones from models with new radiative opacities. *Astron. Astrophys. Suppl. Ser.* **106**, 275–302 (1994).
- Gaudi, B. S. Distinguishing between binary-source and planetary microlensing perturbations. *Astrophys. J.* **506**, 533–539 (1998).
- Glanz, J. Flickers from far-off planets. *Science* **277**, 765 (1997).
- Rhie, S. H. *et al.* On planetary companions to the MACHO-98-BLG-35 microlens star. *Astrophys. J.* (in the press); preprint astro-ph/9905151 at (<http://xxx.lanl.gov>) (1999).
- Jensen, E. L. N. & Mathieu, R. D. Evidence for cleared regions in the disks around pre-main-sequence spectroscopic binaries. *Astron. J.* **114**, 301–316 (1997).
- Bennett, D. P. *et al.* The MACHO Project II: data reduction and analysis of 6 million lightcurves. *BAAS* **183**, 7206 (1993).
- Stetson, P. The center of the core-cusp globular cluster M15: CFHT and HST observations, ALL-FRAME reductions. *Proc. Astron. Soc. Pacif.* **106**, 250–280 (1994).
- Han, C. & Gould, A. The mass spectrum of MACHOs from parallax measurements. *Astrophys. J.* **447**, 53–61 (1995).

Acknowledgements

The authors are members of the MPS Collaboration and the Wise Observatory GMAN Team. We thank the MACHO, OGLE and EROS collaborations for announcements of microlensing events in progress, and the MACHO team for their discovery of MACHO-97-BLG-41. We also thank W. Sutherland for access to his likelihood estimation code. This work was supported in part by the NASA Origins program, the US NSF and a Research Innovation Award for the Research Corporation. Work performed at MSSSO is supported by the Australian Department of Industry, Technology and Regional Development. Work performed at UW is supported in part by the Office of Science and Technology Centers of NSF. Astronomy at Wise Observatory is supported by the Israel Science Foundation.

Correspondence and requests for materials should be addressed to D.P.B. (e-mail: bennett@nd.edu).

The inverse band-structure problem of finding an atomic configuration with given electronic properties

Alberto Franceschetti & Alex Zunger

National Renewable Energy Laboratory, Golden, Colorado 80401, USA

Modern crystal-growth techniques, such as molecular beam epitaxy or metal-organic chemical-vapour deposition, are capable of producing prescribed crystal structures, sometimes even in defiance of equilibrium, bulk thermodynamics. These techniques open up the possibility of exploring different atomic arrangements in search of a configuration that possesses given electronic and optical properties¹. Unfortunately, the number of possible combinations is so vast, and the electronic properties are so sensitive to the details of the crystal structure, that simple trial-and-error methods (such as those used in combinatorial synthesis²) are unlikely to be successful. Here we describe a theoretical method that addresses the problem of finding the atomic configuration of a complex, multi-component system having a target electronic-structure property. As an example, we predict that the configuration of an $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ alloy having the largest optical bandgap is a $(\text{GaAs})_2(\text{AlAs})_1(\text{GaAs})_4(\text{AlAs})_1$ superlattice oriented in the [201] direction.

Conventional electronic-structure theory of solids, clusters and molecules proceeds by first specifying the spatial coordinates of all atomic species involved, and then calculating the ensuing energy levels and wavefunctions:

$$\text{Atomic configuration} \rightarrow \text{Electronic structure} \quad (1)$$

We are interested in the counterpart to this direct approach: the 'inverse approach' of finding the atomic configuration that produces a prescribed electronic structure:

$$\text{Electronic structure} \rightarrow \text{Atomic configuration} \quad (2)$$

In the context of optical properties, such a method would provide answers to questions like "For a given superlattice orientation, what is the layer sequence that has the maximum bandgap (or a pre-assigned bandgap, say 2 eV)?"³. In the context of transport properties, one could ask: "What is the crystal structure whose band structure maximizes Auger carrier multiplication?"⁴. Similar 'inverse problems' can be addressed in the context of vibrational and photonic properties, as well as for molecules and low-dimensional systems.

We describe here a solution to the 'inverse band-structure problem' based on a direct exploration of the space of atomic configurations in search of the configuration possessing given electronic properties. The formidable complexity of this problem is best illustrated by an example. Consider a pseudo-binary substitutional alloy $\text{A}_x\text{B}_{1-x}\text{C}$ of composition x , described by a unit cell of $2N$ lattice sites, such that each of the N cation sites is occupied by an A atom or a B atom. The number of possible atomic configurations is $N_{\text{config}} = N!/((xN)!(N-xN)!)$. For instance, for $x = 0.25$ and $2N = 128$, $N_{\text{config}} \approx 10^{14}$. Each configuration has, in principle, a different electronic structure, and therefore different electronic and optical properties. How many atomic configurations (N_{search}) is it necessary to explore in order to find the configuration that has the target electronic properties? Clearly, if this approach is to be successful N_{search} should be small compared to N_{config} . We search the configuration space using a simulated-annealing algorithm that is able to 'learn the system' relatively quickly ($N_{\text{search}} \approx 10^4$) by retain-

ing only those configurations that are conducive to the target electronic structure. Coupled with a fast, "Order N " solution of the Schrödinger equation, this method allows us to determine the target configuration of systems containing a few hundred atoms.

We consider substitutional systems described by an underlying lattice structure of N sites. An atomic configuration σ is defined as a set of N 'occupation variables' $\{S_1, S_2, \dots, S_N\}$, where S_i denotes the identity of the atom located at lattice site i . The configuration variables $\{S_i\}$ can be restricted to describe a particular experimental growth sequence, for example an $\text{A}_p\text{B}_q\text{A}_r\text{B}_s \dots$ superlattice (where p, q, r, s denote the number of monolayers) oriented in the [001] direction. We then consider a set of electronic-structure properties $P_\alpha(\sigma)$, defined for every configuration σ , and the corresponding target properties P_α^{target} . The subscript α identifies specific properties, such as bandgaps, effective masses or oscillator strengths. The N -variable object function

$$O(\sigma) = \sum_{\alpha} \omega_{\alpha} |P_{\alpha}(\sigma) - P_{\alpha}^{\text{target}}|, \quad (3)$$

where ω_{α} is the weight assigned to the property P_{α} , describes the 'distance' between the electronic structure of the configuration σ and the target electronic structure. This function $O(\sigma)$ is minimized by varying the configuration variables $\sigma = \{S_1, S_2, \dots, S_N\}$, and calculating at each step $P_{\alpha}(\sigma)$ from electronic-structure theory. This approach requires a fast-learning method of sampling the configuration space, and a numerically efficient yet physically accurate method of calculating the electronic structure of a given atomic configuration. Here we describe these methods.

(1) *Exploration of the configuration space.* The simulated-annealing technique is an efficient algorithm for finding the global minimum of a multi-variable, multi-valley function. Given an atomic configuration σ , a trial configuration σ_{trial} is generated by elementary Monte Carlo moves, such as changing the identity of one atom, or swapping the positions of two atoms of different types. The trial configuration σ_{trial} is accepted with a probability distribution

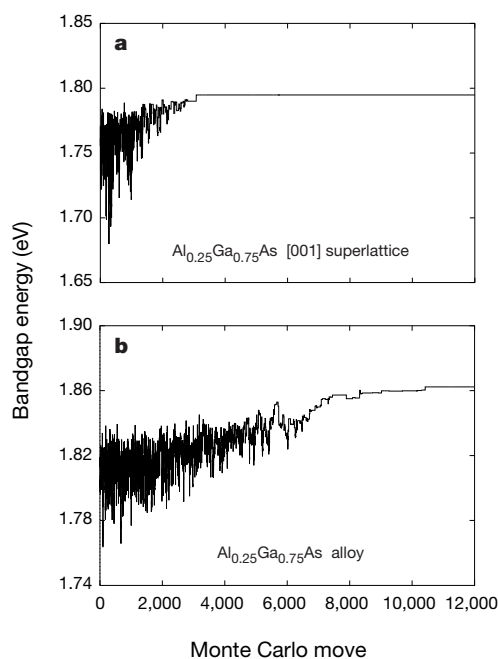


Figure 1 Simulated-annealing search of the maximum-gap configuration. The bandgap is shown as a function of the number of elementary Monte Carlo moves. The simulation cell includes 64 atoms for superlattices oriented in the [001] direction and 128 atoms for alloys. **a**, [001] $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ superlattices; **b**, $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ alloys.

Table 1 Maximum-gap configurations of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{P}$ superlattices oriented in the [001] direction

Concentration x	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	$\text{Ga}_x\text{In}_{1-x}\text{P}$
0.250	(AlAs) ₁ (GaAs) ₃	(GaP) ₁ (InP) ₃
0.333	(AlAs) ₁ (GaAs) ₂	(GaP) ₁ (InP) ₂
0.500	(AlAs) ₂ (GaAs) ₂	(GaP) ₁ (InP) ₁
0.666	(AlAs) ₂ (GaAs) ₁	(GaP) ₂ (InP) ₁
0.750	(AlAs) ₃ (GaAs) ₁	(GaP) ₃ (InP) ₁

$p(\Delta O) = \min[1, \exp(-\Delta O/T)]$ where $\Delta O = O(\sigma_{\text{trial}}) - O(\sigma)$ and T is a fictitious temperature parameter. By slowly decreasing the temperature to zero, the system is allowed to settle into the configuration that minimizes the object function O . We use an exponential temperature profile $T_n = T_0 \exp(-n/\tau)$, where T_0 is the initial temperature and τ is the temperature decay rate. Here n is the index of the simulated-annealing steps; each simulated-annealing step includes several elementary moves. The exponential temperature profile allows us to reduce the acceptance ratio in a nearly linear fashion. If the annealing process is too fast the system can become trapped in a local-minimum configuration. We can detect and avoid local minima by changing the initial configuration and/or the temperature profile.

(2) *Electronic structure of a given atomic configuration.* This is done in three steps:

First, atomic relaxations: For each elementary Monte Carlo move the system is relaxed into the local energy minimum corresponding to the selected configuration (that is, without interchange of atoms). This task is accomplished using a valence-force-field method⁵ whose parameters are fitted to *ab initio* total-energy calculations for various configurations. The outputs are the equilibrium atomic positions $\{\mathbf{R}_n, n = 1 \dots N_{\text{at}}\}$ of the selected configuration.

Second, realistic hamiltonian: it is well known⁶ that the conventional first-principle description of optical properties—the local-density approximation (LDA) to density-functional theory—produces systematic errors in the bandgap, a quantity that we are interested in here. We thus do not use the LDA approach. Instead, the electronic structure of each configuration is described by a semi-empirical pseudopotential hamiltonian:

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + \sum_{n=1}^{N_{\text{at}}} v_n(\mathbf{r} - \mathbf{R}_n) \quad (4)$$

where m is the bare electron mass. The total pseudopotential is given by a linear superposition of local atomic pseudopotentials $v_n(\mathbf{r} - \mathbf{R}_n)$ centred at the atomic positions $\{\mathbf{R}_n, n = 1 \dots N_{\text{at}}\}$. Unlike conventional empirical pseudopotentials⁷, the atomic

pseudopotentials v_n are fitted⁸ not only to observed bulk interband transition energies, but also to (1) effective masses, deformation potentials, and band offsets, and (2) bulk single-particle wavefunctions calculated using density-functional theory in the local-density approximation. Since we are dealing here with bulk-like atomic configurations that are free from defects or surfaces (each cation is coordinated by anions and each anion is coordinated by cations), the hamiltonian of equation (4) should provide an accurate description of all the configurations reached by the Monte Carlo sampling. Such modern semi-empirical pseudopotentials have successfully described the electronic and optical properties of substitutional systems, such as semiconductor alloys and superlattices^{9,10}.

Third, fast diagonalization: Since we are interested in electronic properties that involve band-edge energy levels and wavefunctions, the hamiltonian of equation (4) is diagonalized using a computationally efficient method that focuses on an energy window around the bandgap¹¹. In this approach, the band-edge energies ϵ_i and wavefunctions ψ_i are obtained by solving the ‘folded-spectrum’ equation:

$$[\hat{H} - \epsilon_{\text{ref}}]^2 \psi_i(\mathbf{r}) = (\epsilon_i - \epsilon_{\text{ref}})^2 \psi_i(\mathbf{r}), \quad (5)$$

where ϵ_{ref} is an arbitrary reference energy. The ‘ground state’ of equation (5) coincides with the eigenstate of the hamiltonian in equation (4) whose energy is closest to the reference energy ϵ_{ref} . Therefore, by choosing the reference energy in the bandgap, the band edge states can be readily obtained by minimizing the functional $A[\psi] = \langle \psi | (\hat{H} - \epsilon_{\text{ref}})^2 | \psi \rangle$. For each Monte Carlo move, the wavefunctions of the last accepted configuration are used as initial guesses for the iterative solution of equation (5).

We consider here AlAs/GaAs and GaP/InP substitutional systems. The underlying lattice has the zinc-blende structure, with cation atoms occupying one of the two interpenetrating f.c.c. sublattices and anion atoms occupying the other f.c.c. sublattice. The AlAs/GaAs system has a small lattice mismatch, so we assume that, for every configuration, the atoms occupy their ideal (zinc-blende) positions. The GaP/InP system, on the other hand, has a substantial lattice mismatch ($\sim 7\%$). Thus, for each configuration the atoms (both cations and anions) are allowed to relax into their local equilibrium positions. Note that the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{P}$ random alloys are characterized by a crossover from a direct-gap semiconductor (with both the valence-band maximum and the conduction-band minimum located at the Γ point of the Brillouin zone) to an indirect-gap semiconductor (with the valence-band maximum at the Γ point and the conduction-band minimum near the X point) as a function of the concentration x . Our approach is capable of addressing substitutional systems that display such types of electronic phase transition.

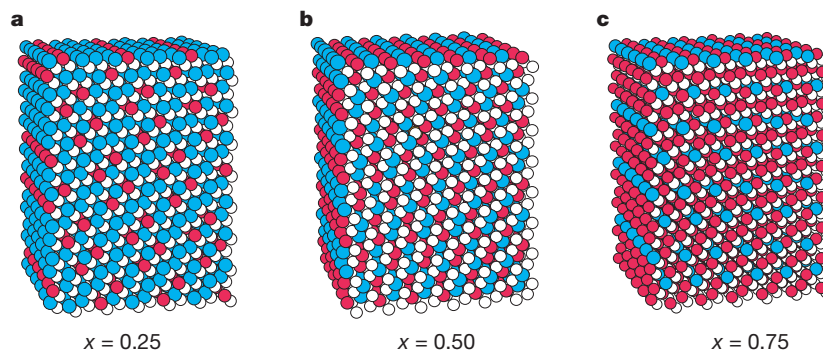


Figure 2 Maximum-gap configurations of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloys for different Al concentrations x . Ga atoms are denoted by blue circles, Al atoms by red circles, and As atoms by white circles. **a**, For $x = 0.25$ the maximum-gap configuration corresponds to a $(\text{GaAs})_2(\text{AlAs})_4$ superlattice in the [201] orientation, and the bandgap is

1.86 eV. **b**, For $x = 0.50$ the maximum-gap configuration is a $(\text{GaAs})_2(\text{AlAs})_2$ superlattice in the [201] direction (chalcopyrite structure), and the bandgap is 2.12 eV. **c**, For $x = 0.75$ the maximum-gap configuration is a $(\text{GaAs})_1(\text{AlAs})_3$ superlattice in the [201] orientation (farnite structure), and the bandgap is 2.18 eV.

The first ‘inverse problem’ we consider is finding the configuration of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{P}$ alloys and superlattices that maximizes the optical bandgap at a given concentration x . Interestingly, the complementary ‘minimum-gap problem’ has a trivial solution for these systems (but not in general, see for example ref. 12), as the configuration that minimizes the bandgap corresponds to the phase-separated alloy for any concentration x . The simulation cell includes 64 atoms for superlattices oriented in the [001] direction and 128 atoms for alloys. While these unit cells may be too small to describe the electronic structure of disordered superlattices or random alloys, we expect that the maximum-gap configuration will be attained by an ordered structure with a relatively small unit cell. Thus, we believe that the structures that maximize the bandgap are fully described by the present simulation cells.

Figure 1 shows representative simulation profiles of the bandgap as a function of the number of elementary Monte Carlo moves. We see that only a few thousand moves are required to reach the configuration having the maximum bandgap ($N_{\text{search}} < 10,000$). This should be compared with the number of possible configurations, which is of the order of $N_{\text{config}} \approx 10^7$ for the 64-atom cell of Fig. 1a and $N_{\text{config}} \approx 10^{14}$ for the 128-atom cell of Fig. 1b. The number of Monte Carlo moves increases superlinearly with the number of atoms in the simulation cell. However, it increases at a much slower rate than the number of configurations.

The maximum-gap configurations of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{P}$ superlattices in the [001] direction are summarized in Table 1. They correspond to the shortest-period superlattices compatible with the assigned cation concentration, with one exception: the $(\text{GaAs})_1(\text{AlAs})_1$ superlattice has a smaller bandgap than the $(\text{GaAs})_2(\text{AlAs})_2$ superlattice, because the folding of the X point of the Brillouin zone into the Γ point causes a strong coupling between the Γ_{1c} and X_{3c} alloy states, pushing the energy of the Γ_{1c} -derived conduction-band minimum of the 1×1 superlattice down.

The maximum-gap configurations of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloys with different Al concentrations x are shown in Fig. 2. We find that the maximum-gap configuration of the $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ alloy (Fig. 2a) corresponds to a $(\text{GaAs})_2(\text{AlAs})_1(\text{GaAs})_4(\text{AlAs})_1$ superlattice in the [201] orientation. Previous ‘cluster expansion’ models³ (describing the bandgap of an arbitrary configuration by a linear combination of energies associated with characteristic ‘figures’, such as pairs or triangles) carried out in the LDA framework had shown that this configuration has the largest bandgap among $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ alloys. However, the value of the bandgap in ref. 3 was affected by the LDA error. For $x = 0.50$ the maximum-gap configuration (Fig. 2b) is a $(\text{GaAs})_2(\text{AlAs})_2$ superlattice in the [201] orientation (chalcopyrite structure). Finally, for $x = 0.75$ the maximum-gap configuration (Fig. 2c) is a $(\text{GaAs})_1(\text{AlAs})_3$ superlattice in the [201] direction (famatinite structure). All the maximum-gap configurations of AlGaAs alloys are predicted to be superlattices in the [201] direction. This result can be understood in terms of folding relationships. For [201] superlattices the X and W points of the bulk Brillouin zone are folded into the Γ point of the superlattice Brillouin zone. However, the X- and W-derived states are weakly coupled to the Γ -derived band-edge states. As a result, [201] superlattices are not affected by the band-gap reduction characteristic of other superlattice orientations. Note that while the $(\text{GaAs})_2(\text{AlAs})_1$ $(\text{GaAs})_4(\text{AlAs})_1$ structure has a direct bandgap (large oscillator strength), the chalcopyrite and famatinite structures have an indirect bandgap (small oscillator strength).

As a second example, we have calculated the configurations of AlAs/GaAs and GaP/InP superlattices in the [001] orientation having a pre-assigned bandgap (1.8 eV for AlAs/GaAs superlattices and 1.9 eV for GaP/InP superlattices) and the largest possible oscillator strength. In these simulations the concentration x is allowed to change; the lattice constant, however, is kept fixed at the average value of the lattice constants of the binary constituents.

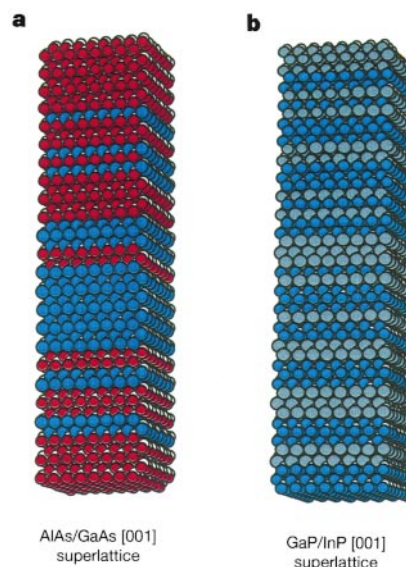


Figure 3 Atomic configurations of superlattices in the [001] orientation having a pre-assigned bandgap and maximum bandgap oscillator strength. The target bandgap is 1.8 eV for AlAs/GaAs superlattices and 1.9 eV for GaP/InP superlattices. Ga atoms are denoted by blue circles, Al atoms by red circles, In atoms by grey circles, and As and P atoms by white circles. **a**, AlAs/GaAs superlattice; **b**, GaP/InP superlattice.

The prescribed bandgap is achieved within a few milli-electronvolts. The atomic configurations having the target electronic properties are shown in Fig. 3. In the case of AlAs/GaAs (Fig. 3a), the ‘core’ of the resulting configuration is a $(\text{GaAs})_2(\text{AlAs})_2(\text{GaAs})_9(\text{AlAs})_2$ $(\text{GaAs})_2$ quantum well embedded in AlAs. The additional Ga layers in the structure do not affect the electronic properties significantly. In the case of GaP/InP (Fig. 3b), the resulting configuration is a superlattice with layer thicknesses ranging from one to three layers.

It is interesting to speculate whether the structures identified via an intensive computer search could have been anticipated based on analytical exploration or physical insight. We suspect that although in some simple cases, such as the short-period superlattices of Table 1, one can rationalize the results after the fact in terms of band-folding relationships, in general the non-intuitive relation between the reciprocal-space band structure and the real-space atomic configuration (equation (2)) makes *a priori* guesses unlikely. In fact, some of the structures we predict are unsuspected on the basis of the normal insights underlying band theory.

Finally, we have applied our method to semiconductor alloys and superlattices, but the same algorithm could be applied to the optimization of the electronic structure of complex molecules and clusters. □

Received 3 May; accepted 20 September 1999.

1. Capasso, F. *Physics of Quantum Electron Devices* (Springer, Berlin, 1990).
2. Wilson, S. R. & Czarnik, A. W. *Combinatorial Chemistry: Synthesis and Application* (Wiley, New York, 1997).
3. Magri, R. & Zunger, A. Real-space description of semiconducting band gaps in substitutional systems. *Phys. Rev. B* **44**, 8672–8684 (1991).
4. Werner, J. H., Kolodinski, S. & Queisser, H. J. Novel optimization principles and efficiency limits for semiconductor solar cells. *Phys. Rev. Lett.* **72**, 3851–3854 (1994).
5. Silverman, A., Zunger, A., Kalish, R. & Adler, J. Atomic-scale structure of disordered $\text{Ga}_{1-x}\text{In}_x\text{P}$ alloys. *Phys. Rev. B* **51**, 10795–10816 (1995).
6. Zunger, A. & Freeman, A. J. Ground- and excited-state properties of LiF in the local-density formalism. *Phys. Rev. B* **16**, 2901–2926 (1977).
7. Chelikowsky, J. R. & Cohen, M. L. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Phys. Rev. B* **14**, 556–582 (1976).
8. Wang, L. W. & Zunger, A. Local-density-derived semiempirical pseudopotentials. *Phys. Rev. B* **51**, 17398–17416 (1995).
9. Mader, K., Wang, L. W. & Zunger, A. Electronic structure of intentionally disordered AlAs/GaAs superlattices. *Phys. Rev. Lett.* **74**, 2555–2558 (1995).

10. Mattila, T., Wei, S.-H. & Zunger, A. Electronic structure of "sequence mutations" in ordered GaInP₂ alloys. *Phys. Rev. Lett.* **83**, 2010–2013 (1999).
11. Wang, L. W. & Zunger, A. Solving Schrödinger's equation around a desired energy: Application to silicon quantum dots. *J. Chem. Phys.* **100**, 2394–2397 (1994).
12. Bellaiche, L., Wei, S. H. & Zunger, A. Localization and percolation in semiconductor alloys: GaAsN vs. GaAsP. *Phys. Rev. B* **54**, 17568–17576 (1996).

Acknowledgements

This work was supported by the US Department of Energy, Office of Science, Division of Materials Science.

Correspondence and requests for materials should be addressed to A.F. (e-mail: afrances@nrel.gov).

Frozen-bed Fennoscandian and Laurentide ice sheets during the Last Glacial Maximum

Johan Kleman* & Clas Hättestrand*

Department of Physical Geography, Stockholm University, S-106 91 Stockholm, Sweden

* The authors contributed equally to this work

The areal extents of the Laurentide and Fennoscandian ice sheets during the Last Glacial Maximum (about 20,000 years ago) are well known¹, but thickness estimates range widely, from high-domed² to thin³, with large implications for our reconstruction of the climate system regarding, for example, Northern Hemisphere atmospheric circulation and global sea levels. This uncertainty stems from difficulties in determining the basal temperatures of the ice sheets and the shear strength of subglacial materials⁴, a knowledge of which would better constrain reconstructions of ice-sheet thickness. Here we show that, in the absence of direct data, the occurrence of ribbed moraines in modern landscapes can be used to determine the former spatial distribution of frozen- and thawed-bed conditions. We argue that ribbed moraines were formed by brittle fracture of subglacial sediments, induced by the excessive stress at the boundary between frozen- and thawed-bed conditions resulting from the across-boundary difference in basal ice velocity. Maps of glacial landforms from aerial photographs of Canada and Scandinavia reveal a concentration of ribbed moraines around the ice-sheet retreat centres of Quebec, Keewatin, Newfoundland and west-central Fennoscandia. Together with the evidence from relict landscapes that mark glacial areas with frozen-bed conditions, the distribution of ribbed moraines on both continents suggest that a large area of the Laurentide and Fennoscandian ice sheets was frozen-based—and therefore high-domed and stable—during the Last Glacial Maximum.

The glaciological factors controlling the location of thermal zones (surface temperature, ice thickness, geothermal flux, strain heating) under ice sheets are well understood⁵, but because of the absence of direct subglacial palaeotemperature records, the actual basal thermal conditions under the Fennoscandian ice sheet (FIS) and Laurentide ice sheet (LIS) have remained elusive. Wide zones of subglacial till (glacially transported sediments) deformation have been invoked in modelling experiments predicting low ice-sheet profiles³, but without data on the phase state (frozen or thawed) of the sediments in deformable bed areas under glacial conditions, the validity of these models has been difficult to verify. The problem is that the rheology of subglacial till is extremely sensitive to the phase state of the interstitial water; frozen subglacial till is much stronger than ice, whereas thawed till under high water pressures can be deformed by overriding ice⁶. We focus here on the landform record

interpreted to result from the phase-state control on soil strength (and thereby landform-building processes) and perform an inversion of the record: that is, we use the spatial distribution of diagnostic landforms to gain insight into the former subglacial phase- and temperature-regime under the FIS and LIS.

The two main subglacial landform groups resulting from reshaping of subglacial sediments are: (1) drumlins and flutings (flow-parallel streamlined till ridges, 0.1–20 km in length), created by particle-by-particle entrainment and lodgement processes or plastic deformation of till masses⁷, and (2) ribbed moraines or Rogen moraines (fields of till ridges, 0.1–1 km in length, formed transverse to ice flow), with a debated mode of formation⁸. Drumlins and flutings cover 90% of the terrestrial parts of the LIS^{9,10} and FIS areas, but ribbed moraines less than 10%, with a spatial distribution that is extremely selective when compared to the lineation distribution. There is a third important group of subglacial landscapes: those which have been left essentially unmodified by the last ice sheet due to sustained frozen-bed conditions and absence of basal sliding¹¹.

Previous formation hypotheses for ribbed moraine suggested compressive flow (caused by topography, that is, by bedrock depressions)¹², or the type and amount of basal debris load¹³ as the primary controls on its formation. As hilly topography and coarse-grained tills are widespread in areas of Precambrian basement rocks (shown yellow in Fig. 1a and b), these hypotheses imply that ribbed moraines would occur geographically dispersed, wherever local topographic or substratum conditions were favourable. The concentration of ribbed moraines in the four retreat centres of Quebec, Keewatin, Newfoundland and west-central Fennoscandia (Fig. 1a and b), and their absence in other areas of hilly relief and coarse-grained tills, indicates that some other type of primary control on their distribution exists.

Ribbed moraines are conspicuously lacking in the southern parts of both ice-sheet areas where thawed-bed conditions prevailed after the Last Glacial Maximum (LGM), and it thus appears unlikely that ribbed moraine formed under completely thawed-bed conditions. The concentric arrangement of ribbed moraines around late-glacial retreat centres, and their affinity to frozen-bed areas, indicates that the primary control is a time-dependent basal thermal evolution. The stratigraphical and structural composition of ribbed moraine is highly variable (from laminated silt to shattered bedrock), and generally follows local variations in the till composition^{8,14}. Hence, ribbed moraines cannot be linked to any specific depositional facies. Rather, the deposition of the sediments in ribbed moraine ridges appears to be genetically unrelated to the actual landform-shaping process^{8,15}, predating it. Morphological evidence (Fig. 2a; detailed till block outline matching, grating patterns, strike-slip faulting of till slabs, rotation of discrete till blocks) indicate that individual ridges were once part of a coherent drift sheet, and we thus infer that formation of ribbed moraine occurred by brittle fracture of drift sheets. Such brittle behaviour cannot occur in unfrozen drift (due to its low cohesive strength), and we therefore argue for a primary thermal control on ribbed moraine formation, by fracturing of frozen drift sheets during the transition from frozen to thawed conditions in an extensional basal regime (Fig. 3). Extreme stress concentrations develop at frozen–thawed boundaries¹⁶, because of the abrupt increase in basal ice flow velocity, and we infer brittle fracture of drift sheets to have occurred at migrating frozen–thawed boundaries. The often observed lack of material between ribbed moraine ridges¹⁵ is not compatible with alternative formation processes of glaciotectionic stacking and sediment thickening¹². The typical 150–600 m spacing of individual ribbed moraine ridges contrasts sharply with the observed tightly folded "wrinkled-carpet" morphology of subglacial compressional ridges¹⁷.

Relict landscapes are defined by ground surfaces and landforms essentially unmodified during overriding by the last ice sheet¹⁸ (Fig. 2b and c). In these landscapes glacial meltwater traces from the last deglaciation (that is, meltwater channels and ice dammed

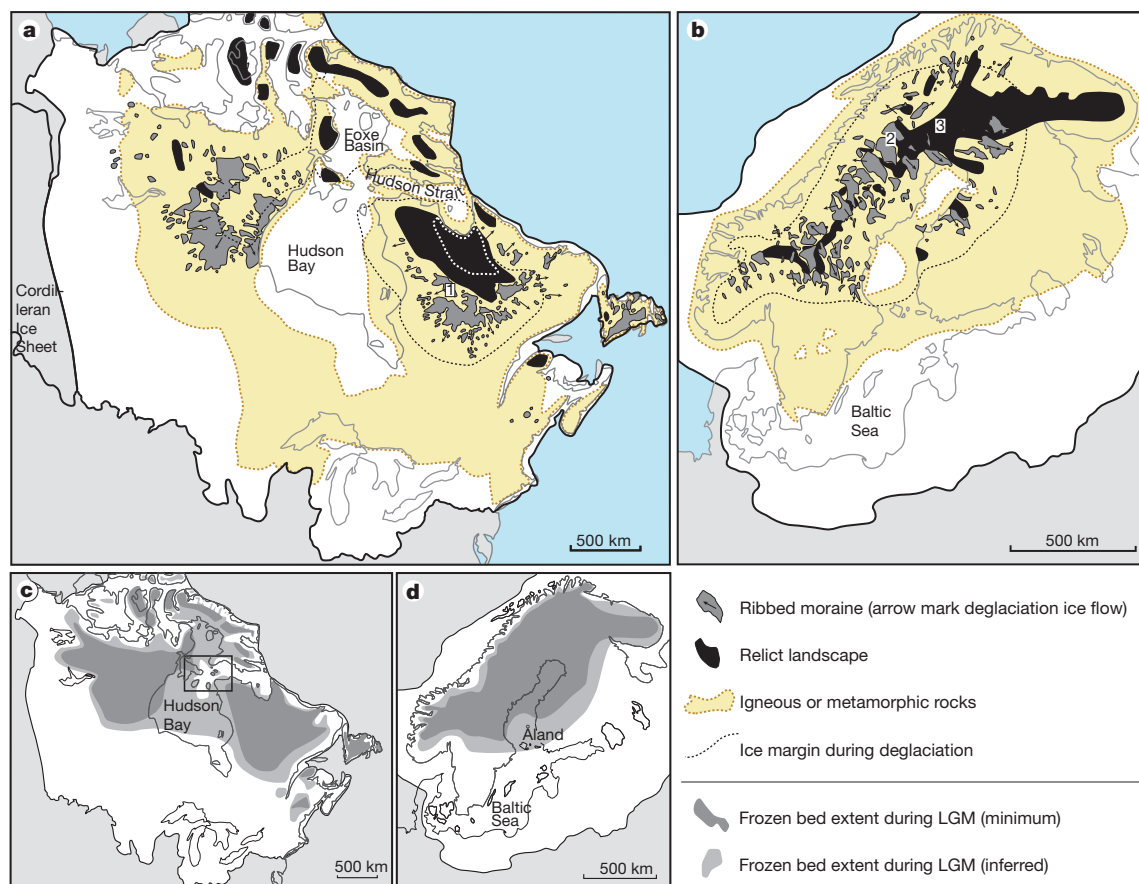


Figure 1 Frozen-bed extent reconstructed from ribbed moraines and relict landscapes. Ribbed moraines (grey) and relict landscapes (black) are shown in the LIS^{10,13,19–22} (a) and FIS^{11,15,23,24,28,31} (b) areas. Arrows mark deglacial (final) ice flow directions, the heavy black line shows the LGM ice-sheet margins, and the thin broken black line shows ice margins at 8,000 radiocarbon years before present (¹⁴C yr BP; LIS) and 9,500 ¹⁴C yr BP (FIS). Areas with exposed igneous or metamorphic rocks, commonly overlaid by discontinuous coarse-grained tills, are shown in yellow. Areas of sedimentary cover rocks, mostly overlain by fine-grained diamictons, are shown in white. Relict landscapes¹⁸ mark the sites of sustained frozen-bed conditions, and ribbed moraine areas mark areas of frozen- to thawed-bed conversion¹⁵. Ribbed moraines exist also south of Ungava Bay (within white dashed line), but are considered to be unrelated to the last deglaciation²¹. The digit '1'

shows location of Fig. 2a; '2' shows location of Fig. 2b; '3' shows location of Fig. 2c.

c, d, Reconstruction of frozen-bed extent under the LIS and FIS during the LGM. Dark shading represents the minimum extent of zones dominated by frozen-bed conditions, directly based on geological evidence (ribbed moraine and relict landscapes). Lighter shading represents probable LGM frozen-bed extent, with allowance made for deglacial inward-transgression of thawed/frozen bed boundaries. Minor areas of thawed bed may have existed in frozen-bed zones but would have had negligible effect on overall ice-sheet dynamics because basal ice flow in a mosaic of frozen and thawed bed patches is largely controlled by the frozen patches²⁹. The boxed area indicates an area critical for discharge of interior Laurentide ice. Frozen-bed sticky spots may have existed on islands at the western end of the Hudson Strait.

lake features) often overprint land surfaces with directionally unaligned glacial landforms, interstadial periglacial features, or pre-glacial weathering remnants. Relict landscapes are particularly well documented in Fennoscandia¹¹ but are also known from many smaller areas in the northern and eastern sectors of the Laurentide area^{19,20}. We regard relict landscapes as frozen-bed markers because it is unlikely that delicate pre-existing landforms, such as metre-scale periglacial boulder polygons¹⁷ (Fig. 2b) and tors (Fig. 2c) could survive thawed-bed conditions and basal sliding for any length of time.

We mapped glacial landforms in one of the two LIS core areas, Quebec-Labrador, in aerial photographs at a scale of 1:60,000, and established the approximate extent of relict landscapes and melt-water landforms (strictly marginal channels, but no eskers from the final glaciation) indicative of impermeable and cold ice during deglaciation²¹. For the more peripheral parts of this core area we used existing maps¹⁰ and aerial-photograph interpretation. For the ribbed moraine distribution in the other LIS core area, Keewatin, we used existing map information^{10,13}. For identifying relict landscapes in the northern and eastern part of the LIS area, we used regional descriptions of glacial geology and geomorphology^{19,20,22}. To establish the spatial distribution of ribbed moraine in the FIS area, we mapped Sweden, which comprises most of the Fennoscandian

occurrences, from high-altitude aerial photographs¹⁴ at a scale of 1:150,000. For the more peripheral parts of the FIS area we used existing map data on the spatial distribution of ribbed moraine^{23,24}. We mapped relict surfaces in the Swedish part of the Scandinavian mountain chain, using colour infrared aerial photographs at a scale of 1:60,000, and established the spatial extent of relict landscapes in other parts of Fennoscandia by analysis of existing map data on glacial geomorphology and surficial deposits. As to data quality: we regard the distribution of ribbed moraine as well established in both the LIS and FIS areas. The extent of relict landscapes in the FIS area is reasonably well known, but in the LIS area there is an uneven coverage, with research into identifying relict landscapes mainly having been focused at the northern and eastern parts.

In two regions, Quebec-Labrador and west-central Fennoscandia, ribbed moraine surrounds relict landscapes indicating frozen-bed conditions during deglaciation. The Keewatin area differs from Fennoscandia and Quebec-Labrador in that there is no surviving central area of relict landscape. Thus, lineation swarms¹⁰ (which are superimposed on the ribbed moraine) extend all the way to the position of the former ice divide. In the Keewatin sector, therefore, ribbed moraine is the main indicator of the extent of the former frozen-bed core area.

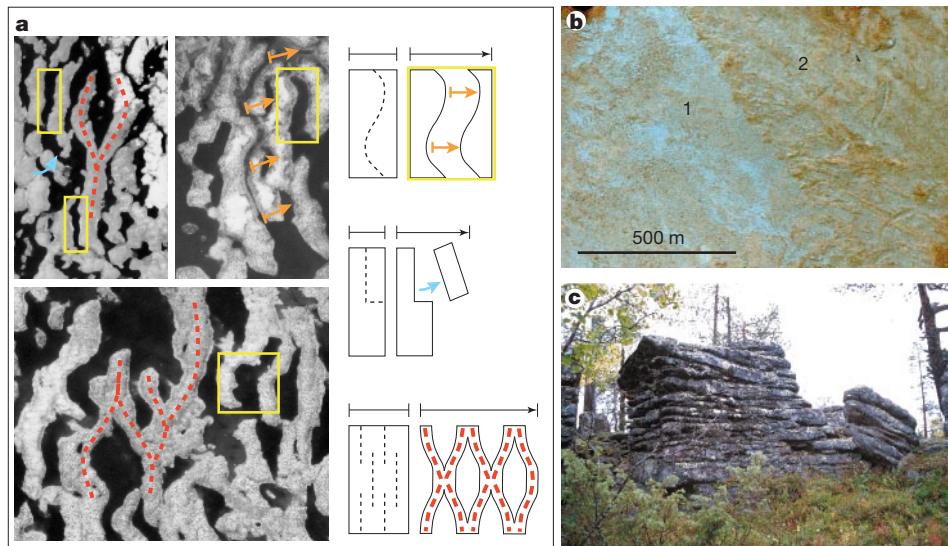


Figure 2 Landforms indicating former frozen-bed conditions. **a**, Vertical aerial photographs of ribbed moraines in central Quebec-Labrador. Ice flow direction in all photographs is from left to right. Three types of dislocation features show that the till sheet underwent extension and fracturing; grating splays (broken red lines), slab detachment and rotation (curved blue arrow), and detailed ridge outline matching (yellow boxes). An embryonic split between two till slabs is shown by orange arrows. Grating splays can only occur in material with considerable tensile strength and thus indicate that fracturing occurred when the till was in a frozen state. **b**, Relict surface of interstadial patterned ground (sorted boulder polygons), labelled '1', truncated by glacial fluting ('2') formed by

basal sliding or till deformation in a thawed-bed zone during deglaciation. This image is a colour infrared photo of Sjuva Hill (800 m elevation) at 66° 37' N, 18° 03' E, northern Sweden. **c**, Granite tor (weathering remnant) at Paljaslaki (390 m elevation), northeastern Sweden. The height of the tor is 2 m. Tors are typical landforms in unglaciated regions, indicating long exposure to subaerial weathering and erosion processes, and where found in previously glaciated areas, such as in northern Sweden, they are indicative of negligible glacial erosion due to frozen-bed conditions²⁰. The locations of these photographs are shown in Fig. 1.

Our evidence suggests that the LIS had wide terrestrial frozen-bed zones (Fig. 1c), similar to the one in Fennoscandia (Fig. 1d), implying high and stable domes over Keewatin and Labrador. The morphological evidence for frozen-bed core areas provides convergence with glaciological theory⁵ that predicts terrestrial core areas, especially at higher elevation, as likely sites for frozen-bed conditions. The results also converge with thermal modelling experiments suggesting frozen-bed conditions in central ice sheet areas, possibly slowly consumed by basal melting late in the glacial cycle²⁵, and time-dependent three-dimensional modelling of Northern Hemisphere ice sheets which suggests frozen-bed core areas²⁶. Oxygen-isotope analyses²⁷ of Pleistocene ice in the Penny ice cap, Baffin island, indicate that the ice is not of local origin, but originated at a high-elevation source over Hudson Bay and/or Foxe basin. Frozen bed conditions in Hudson Bay during the LGM were inferred²⁷.

Because of the cumulative nature of subglacial reshaping, the present extent of ribbed moraines and relict landscapes most probably represents only a part of the area that at any given time had a frozen bed. Hence, we have shown (Fig. 1c and d) two alternative frozen-bed extents during the LGM; the smallest we consider the data to allow, and a more tentative extent where allowance has been made for the likely destruction of morphological evidence in peripheral sliding zones that transgressed inwards over former frozen-bed areas during final deglaciation. As shown in Fig. 4, the duration of (final) wet-based flow was longer in the more peripheral parts of the ice sheet, and major (possibly complete), reshaping and destruction is therefore likely to have affected the oldest and most peripheral ribbed moraines. The shrinkage of the frozen-bed zone mainly occurred by slow retreat of the boundary between the frozen and the thawed bed (Fig. 4), but was locally enhanced by rapid headward propagation of discrete radial wet-bed zones during development of late glacial ice streams²⁸.

For the overall ice sheet geometry in the LIS area, conditions at the junction between Hudson Bay, the Hudson Strait and Foxe basin must have been crucial. These areas are mainly underlain by

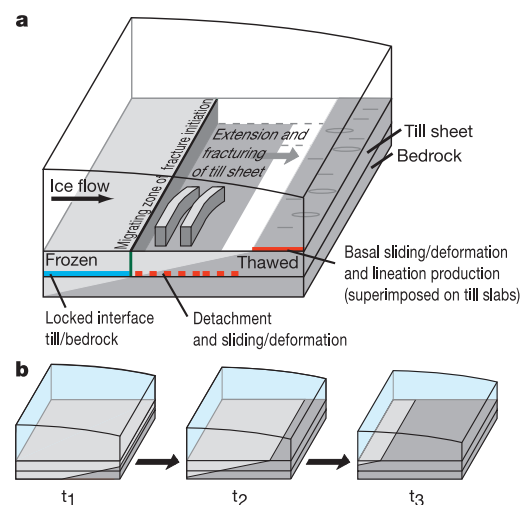


Figure 3 Formative conditions for ribbed moraine. **a**, Block diagram of the three-layer sandwich ice-till-bedrock, where the two interfaces can be either locked (frozen), or unlocked (thawed). During thawing of a glacier bed from underneath, a phase-change surface (separating frozen and thawed material) intersects the glacier bed at low angle, sloping upward in the ice-flow direction. Where the phase-change surface intersects the till-bedrock interface, the basal ice flow velocity will increase as the bed thaws, causing a zone of strong longitudinal tensile stresses. Fracturing of the still frozen till sheet is likely to occur immediately above the detachment zone (broken red line), down-ice of the locked till-bedrock interface (blue line), creating ribbed moraine. Further downglacier, where the thawing front reaches the till-ice interface (red line), basal sliding, till deformation and lineation production can take place. **b**, The sequence **t**₁–**t**₃ shows ice sheet thinning and retreat of the frozen/thawed bed boundary during deglaciation. We interpret the thawed/frozen bed boundary to have transgressed inwards during final decay of mid-latitude ice sheets, causing a narrow fracture-inducing zone of tensile stresses to transgress over till sheets and creating ribbed moraine successively up-ice. Ribbed moraine probably formed as a consequence of non-equilibrium conditions unique to ice-sheet decay phases (ice-marginal retreat and slow subglacial thawing).

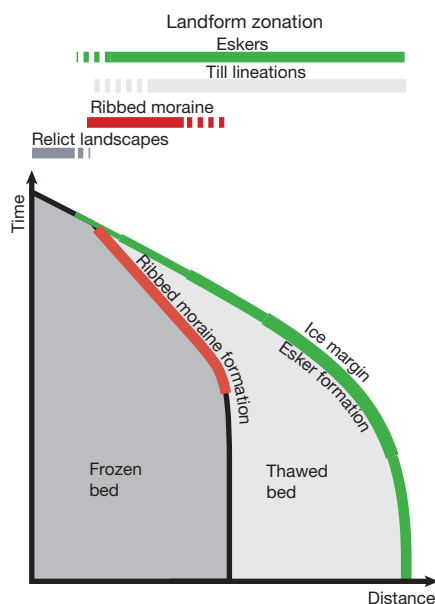


Figure 4 Time–space domains of glacial landform formation. Distance is shown along a transect from dispersal centre to ice margin. Time is shown from the LGM to the time of deglaciation at the final dispersal centre. Bars at the top of the figure show along-transect location of landform types and relict landscapes.

sedimentary cover rocks, the main source of fine-grained sea-floor diamictos in the region. Whether or not these diamictos were deformable (and therefore conducive to low ice-sheet profiles) was a function of ice-sheet basal temperatures in this region, and cannot be assumed *a priori*. Horsts of Precambrian rock (Nottingham, Salisbury and Mill islands, and the northern part of Southampton island) stand 200–600 m above the adjacent sea floor and may have acted as frozen-bed ‘sticky spots’²⁹ in a region elsewhere characterized by thawed-bed conditions, thereby substantially retarding ice outflow through the Hudson Strait and raising the surface elevations over Hudson Bay. This is a situation analogous to the Åland islands in the FIS area, which modelling experiments indicate to have been frozen-bed sticky spots hindering headward propagation of the Baltic ice stream into the central Fennoscandian ice sheet area³⁰.

Our results show that a time-dependent basal thermal evolution was the primary control on bed traction in each of the FIS and LIS areas. Large fractions of the LIS and FIS beds were frozen, but thawed during the post-LGM stages. This created the conspicuous landform ribbed moraine, and efficiently lowered ice-sheet profiles by allowing basal sliding and deformation of formerly frozen and undeformable till. □

Received 15 February; accepted 13 September 1999.

- Denton, G. H. & Hughes, T. J. (eds) *The Last Great Ice Sheets* (Wiley-Interscience, New York, 1981).
- Budd, W. F. & Smith, I. N. in *Sea Level, Ice, and Climatic Change* (ed. Allison, I.) 369–409 (Publ. 131, International Association of Hydrological Sciences, Washington DC, 1981).
- Licciardi, J. M., Clark, P. U., Jenson, J. W. & MacAyeal, D. R. Deglaciation of a soft-bedded Laurentide Ice Sheet. *Quat. Sci. Rev.* **17**, 427–448 (1998).
- MacAyeal, D. R. Binge/purge oscillations of the Laurentide Ice Sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography* **8**, 775–784 (1993).
- Hooke, R. LeB. Basal temperatures in polar ice sheets – a qualitative review. *Quat. Res.* **7**, 1–13 (1977).
- Hughes, T. J. in *Proc. 2nd Int. Conf. on Permafrost* 213–223 (National Academy of Sciences, Washington DC, 1973).
- Boulton, G. S. in *Drumlin Symposium* (eds Menzies, J. & Rose, J.) 25–80 (Balkema, Rotterdam, 1987).
- Lundqvist, J. Rogen moraine – an example of two-step formation of glacial landscapes. *Sedim. Geol.* **111**, 27–40 (1997).
- Boulton, G. S. & Clark, C. D. A highly mobile Laurentide Ice Sheet revealed by satellite images of glacial lineations. *Nature* **346**, 813–817 (1990).
- Prest, V. K., Grant, D. R. & Rampton, V. N. *Glacial Map of Canada* (Geological Survey of Canada, Ottawa, 1969).
- Lagerbäck, R. & Robertsson, A. -M. Kettle holes – Stratigraphical archives for Weichselian geology and palaeoenvironment in northernmost Sweden. *Boreas* **17**, 439–468 (1988).

- Bouchard, M. A. Subglacial landforms and deposits in central and northern Quebec, Canada, with emphasis on Rogen moraines. *Sedim. Geol.* **62**, 293–308 (1989).
- Aylsworth, J. M. & Shilts, W. W. Glacial features around the Keewatin ice divide: Districts of Mackenzie and Keewatin. *Geol. Surv. Can. Pap.* **88-24**, 1–21 (1989).
- Hättestrand, C. Ribbed moraines in Sweden – distribution pattern and paleogeological implications. *Sedim. Geol.* **111**, 41–56 (1997).
- Hättestrand, C. & Kleman, J. Ribbed moraine formation. *Quat. Sci. Rev.* **18**, 43–61 (1999).
- Hutter, K. & Olunloyo, V. O. S. Basal stress concentrations due to abrupt changes in boundary conditions: a cause for high till concentration at the bottom of a glacier. *Ann. Glaciol.* **2**, 29–33 (1981).
- Kleman, J. & Borgström, I. Glacial land forms indicative of a partly frozen bed. *J. Glaciol.* **40**, 255–264 (1994).
- Kleman, J. Preservation of landforms under ice sheets and ice caps. *Geomorphology* **9**, 19–32 (1994).
- Dyke, A. S. Landscapes of cold-centred Late Wisconsinan ice caps, Arctic Canada. *Prog. Phys. Geogr.* **17**, 223–247 (1993).
- Sugden, D. E. & Watts, S. H. Tors, felsenmeer and glaciation in northern Cumberland Peninsula, Baffin Island. *Can. J. Earth Sci.* **14**, 2817–2823 (1977).
- Kleman, J., Borgström, I. & Hättestrand, C. Evidence for a relict glacial landscape in Quebec-Labrador. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **111**, 217–228 (1994).
- Dyke, A. S., Morris, T. F., Green, D. E. C. & England, J. *Quaternary Geology of Prince of Wales Island Arctic Canada* (Mem. 433, Geological Survey of Canada, Ottawa, 1992).
- Nordkalott project *Map of Quaternary Geology Sheet 2, Glacial Geomorphology, Northern Fennoscandia, 1:1 mill.* (Geological Surveys of Finland, Norway and Sweden, Geological Survey of Sweden, Uppsala, 1986).
- Sollid, J. L. & Torp, B. *Nasjonaltatlas for Norge, Glacialgeologisk kart over Norge, 1:1 mill.* (Geografisk Institutt, Oslo, 1984).
- Heine, J. T. & McTigue, D. F. A case for cold-based continental ice sheets—A transient thermal model. *J. Glaciol.* **42**, 37–42 (1996).
- Huybrechts, P. & T'siobbel, S. Thermomechanical modeling of Northern Hemisphere ice sheets with a two-level mass-balance parametrization. *Ann. Glaciol.* **21**, 111–116 (1995).
- Fisher, D. A. *et al.* Penny ice cap cores, Baffin Island, Canada, and the Wisconsin fore dome connection: Two states of Hudson Bay ice cover. *Science* **279**, 692–695 (1998).
- Kleman, J., Hättestrand, C., Borgström, I. & Stroeven, A. Fennoscandian palaeoglaciology reconstructed using a glacial geology inversion model. *J. Glaciol.* **43**, 283–299 (1997).
- Hughes, T. J. Can ice sheets trigger abrupt climatic change? *Arct. Alp. Res.* **28**, 448–465 (1996).
- Holmlund, P. & Fastook, J. A time-dependent glaciological model of the Weichselian Ice Sheet. *Quat. Int.* **27**, 53–58 (1995).
- Niemelä, J., Ekman, I. & Lukashov, A. (eds) *Quaternary Deposits of Finland and Northwestern part of Russian Federation and their Resources 1:1 mill* (Geological Survey of Finland, and Institute of Geology, Karelian Science Centre of the Russian Academy of Sciences, Helsinki, 1993).

Acknowledgements

This work was made possible through grants from the Swedish Natural Science Research Council and Carl Mannerfelts fund. We thank R. LeB. Hooke for comments on the manuscript.

Correspondence and requests for materials should be addressed to J.K. (e-mail: kleman@natgeo.su.se).

Limits on differential rotation of the inner core from an analysis of the Earth's free oscillations

Gabi Laske & Guy Masters

Institute of Geophysics and Planetary Physics, University of California San Diego, 9500 Gilman Dr, La Jolla, California 92093-0225, USA

Differential rotation of the Earth's inner core has been inferred by several seismic ‘body-wave’ studies^{1–6} which indicate that the inner core is rotating at a rate between 0.2° and 3° per year faster than the Earth's crust and mantle. The wide range in inferred rotation rate is thought to be caused by the sensitivity of body-wave studies to local complexities in inner-core structure^{3,7}. Free-oscillation ‘splitting functions’, on the other hand, are insensitive to local structure and therefore have the potential to estimate differential rotation more accurately. A previous free-oscillation study⁸, however, was equivocal in its conclusions because of the relatively poor quality and coverage of the long-period digital data available 20 years ago. Here we use a method for analysing free oscillations⁹ which is insensitive to

earthquake source, location and mechanism to constrain this differential rotation. We find that inner-core differential rotation is essentially zero over the past 20 years (to within $\pm 0.2^\circ$ per year), implying that the inner core is probably gravitationally locked to the Earth's mantle¹⁰.

It has long been known that free oscillations which sample the inner core are strongly split¹¹ by a structure which is dominantly axisymmetric and mimics the effect of an excess ellipticity of the Earth. In 1986, workers at Harvard^{12,13} inferred that anisotropy of the inner core was the main reason for the anomalous splitting and this same anisotropy could explain some features of the travel times of body waves which sample this region¹⁴. This explanation has become generally accepted by seismologists though agreement on the physical cause for the phenomenon has been lacking^{15–18}. According to the accepted story, the inner core behaves roughly like a single anisotropic crystal with a seismically fast symmetry axis closely aligned with the rotation axis of the Earth. A small departure of the fast symmetry axis from the rotation axis has been inferred by several groups^{19–22}; this departure implies that any differential rotation of the inner core would result in temporal variations of the travel times of body waves emanating from a particular source region and recorded by a particular receiver. Such temporal variations have been inferred to exist for some paths^{1,3} but not for others⁵. Furthermore, detailed studies of body-wave arrival times using arrays of detectors have shown that strong local lateral variations in anisotropy in the inner core exist, making it much more difficult to infer a differential rotation rate for the inner core.

Free oscillation splitting functions provide a much better tool for looking at the differential rotation rate of the inner core. Free oscillations are natural low-pass filters of three-dimensional structure, so it is efficient to use a spherical harmonic expansion of structure in mode studies. By this, we mean that structure is expanded in the form

$$\delta \mathbf{m}(r, \theta, \phi) = \sum_{s,t} \delta \mathbf{m}_s^t(r) Y_s^t(\theta, \phi)$$

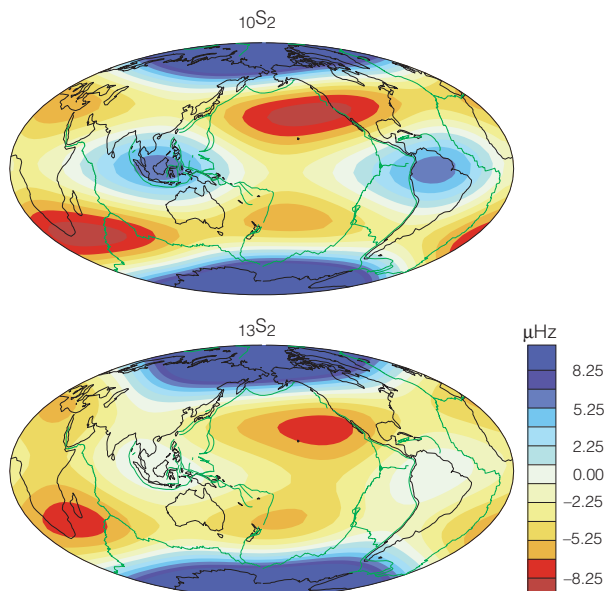


Figure 1 Splitting functions for the two modes $10S_2$ and $13S_2$. The effects of rotation and ellipticity have been removed before the inversion for structure coefficients (equation (2)). The splitting functions are very robust and the non-zonal patterns are very well aligned.

where $Y_s^t = X_s^t(\theta)e^{im\phi}$ is a spherical harmonic of harmonic degree s and azimuthal order number t ; r , θ , ϕ are radius, co-latitude and longitude. Here, $\delta \mathbf{m}$ refers to lateral variations in density and seismic velocities. An isolated mode of harmonic degree s is sensitive to even-order structure only up to harmonic degree $2s$. Thus, modes of harmonic degree 1 are sensitive only to structure of degree 2. Unfortunately, it seems that structure of degree 2 in the inner core is almost completely zonal (azimuthal order = 0) as sensed by modes of harmonic degree 1, so such modes are insensitive to any rotation about the pole. Fortunately, other modes of harmonic degrees greater than 1 appear to be sensitive indicators of inner-core rotation.

We have recently introduced a new technique for analysing the sensitivity of a free oscillation to three-dimensional structure⁹, which is independent of the source mechanism which excites the free oscillation. Briefly, for a particular earthquake and a mode of harmonic degree l , there are $2l + 1$ linear combinations of recordings $\mathbf{b}(t)$ which we call 'receiver strips' which satisfy the following autoregressive relationship:

$$\mathbf{b}(t + \delta t) = P(\delta t)\mathbf{b}(t) \quad (1)$$

Here the matrix $P(\delta t) = \exp[i\delta t(H + I\bar{\omega})]$, $\bar{\omega}$ is the degenerate frequency of the mode, I is the identity matrix and H is the splitting matrix given by^{23,24} (after corrections for the rotation and ellipticity of the Earth)

$$H_{mn} = \sum_{st} \gamma_{mn}^{st} c_s^t \quad (2)$$

where γ represents geometrical factors which can easily be computed, and

$$c_s^t = \int_0^a M_s(\mathbf{r}) \cdot \delta \mathbf{m}_s^t(r) r^2 dr$$

We note that each c_s^t is proportional to a depth-weighted average of a $\delta \mathbf{m}_s^t(r)$ where the weighting kernel, $M_s(r)$, varies from mode to mode. Hence, it is convenient to visualize the geographical distribution of structure sensed by a mode by forming the

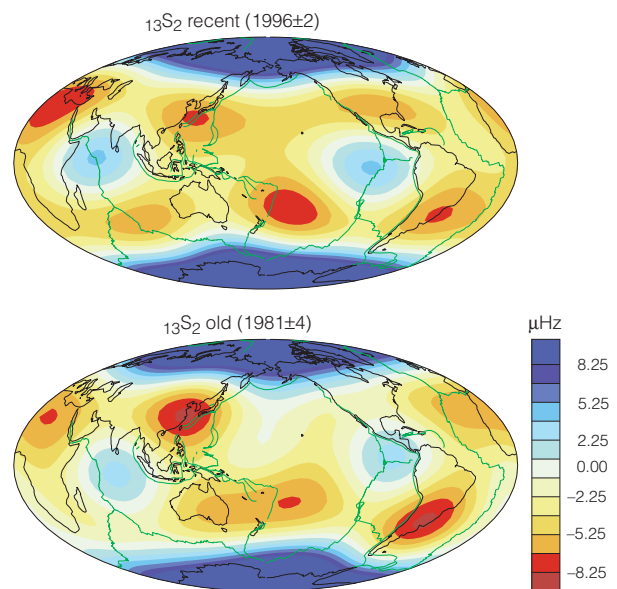


Figure 2 Splitting functions for mode $13S_2$. Panel **a** uses recent events from the years 1996 ± 2 , and panel **b** uses earlier events from the years 1981 ± 4 . The effects of rotation and ellipticity have been removed before the inversion for structure coefficients, and the contribution of the mantle³¹ has been removed. The anomalies shown are due to structure in the inner core. A westward rotation of the lower map about the vertical axis of 4.5° optimally aligns it with the upper map. Provided that our mantle correction is accurate, this comparison implies a westward inner-core rotation of 0.3° per year.

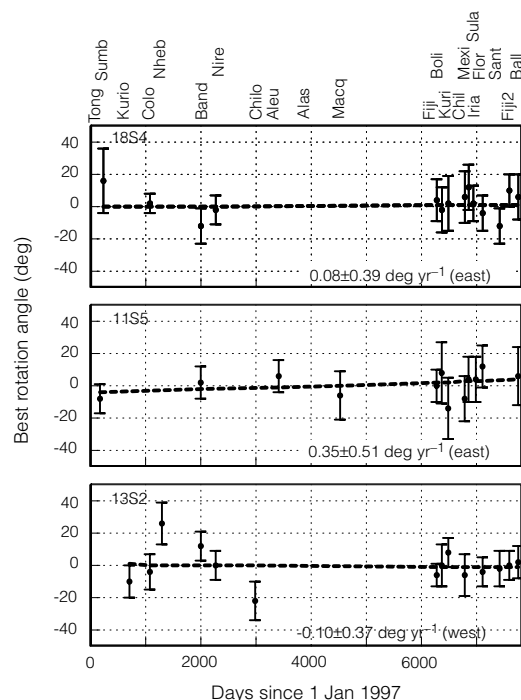


Figure 3 Time-dependent inner-core rotation angles. The angles, shown for the three modes $_{18}S_4$, $_{11}S_5$ and $_{13}S_2$, give the best fit to the receiver strips (see text) of individual events. The dashed lines fit the data in a least-squares sense. Their slopes translate to the annual rotation rates given in the lower right corner of each diagram. Positive rotation rates imply a superrotation of the inner core.

splitting function²⁵:

$$f(\theta, \phi) = \sum_{s,t} c_s^t Y_s^t(\theta, \phi).$$

Our technique uses a linear inverse to estimate P from equation (1) by simultaneously fitting the receiver strips from many earthquakes. Matrix H can easily be recovered from P . We then solve another linear inverse problem to obtain the structure coefficients (equation (2)) and so form splitting functions. Splitting functions for two harmonic degree 2 modes which are sensitive to the inner core are shown in Fig. 1. These splitting functions have been derived using data from 11 large earthquakes which occurred between 1994 and 1998 (Table 1), and have been constructed using about 1,000 vertical component recordings. The measured splitting functions are exceptionally robust and exhibit non-zonal patterns which, interestingly, coincide geographically to better than a degree for the two modes. We take this as a measure of the precision of splitting functions that can be achieved using modern data.

For a mode sensitive to the inner core, a differential rotation will cause the non-zonal structure coefficients to change with time so the splitting function will change with time. We illustrate this idea with the mode $_{13}S_2$ (Fig. 2). We compare the splitting function derived using our technique and recent data (top panel) with the splitting function obtained employing the traditional, nonlinear technique of iterative spectral fitting^{26,27} applied to data covering the timespan 1977–85^{27,28}. We compare the splitting functions after correction for structure in the mantle which we do not expect to change with time. This mantle contribution has been estimated using several different models of mantle structure^{29–31} and appears to be robustly determined. However, it contributes substantially to the patterns shown in Fig. 1 and, in fact, the mantle can explain all the non-zonal signals we see for modes of harmonic degree 1 (for example, $_{8}S_1$, $_{13}S_1$), making these modes useless for determining inner-core rotation.

If we solve for an inner-core rotation that best fits the two mantle-corrected non-zonal patterns for $_{13}S_2$, we find that the best answer

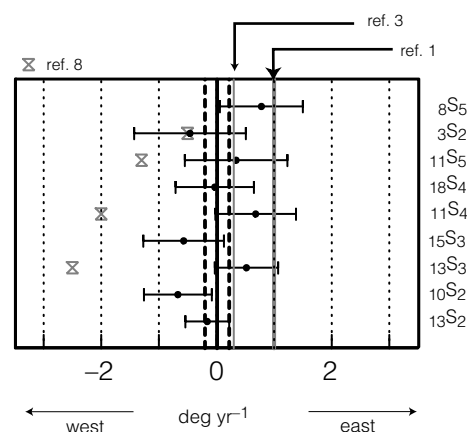


Figure 4 Inner-core rotation rates obtained for nine core-sensitive modes. Also shown is the mean rotation rate, $0.01 \pm 0.21 \text{ deg yr}^{-1}$, obtained from averaging over all modes. Dashed lines mark the error bars of the average. The hour-glass symbols mark the results of a recent mode study⁸, and two grey vertical lines stand for the superrotations found in recent body-wave studies^{1,3}. Our data are marginally consistent with a superrotation of 0.3° per year, though a superrotation of 1° per year is clearly inconsistent for most of the modes. See Table 1 for event abbreviations.

for the total pattern is an annual inner-core rotation of about 0.3° westward, though the weak non-zonal degree 2 part appears to go eastward while the stronger degree 4 part goes westward. Of course, this is physically impossible if the inner core is moving as a rigid body. This result is most likely to be due to errors in the older splitting function which is constrained by much fewer high-quality data than the modern one.

A better way of testing for inner-core rotation is by testing the hypothesis that the inner core is differentially rotating about the rotation axis. We assume that the modern data accurately constrain the current splitting function of the mode, and that we can properly correct for structure in the mantle. The corrected splitting function by assumption reflects only inner-core structure, which can now be rotated about the rotation axis using an assumed rotation rate. We then add the mantle contribution back in, and we have a synthetic H from which we compute a synthetic P . This P is used in equation (1) to test if the assumed rotation rate provides a good fit to the receiver strips, **b**, for any event. The data have to meet certain criteria to be considered in the hypothesis test. For example, the fit to the receiver strips with any non-zero inner-core rotation must be better than that using no rotation, and the receiver strips should be fitted close to their error bars. Not all events excite all modes, so we restrict

Table 1 Earthquakes used in this study

Event name	Year.day	Depth (km)	Moment (10^{25} Nm)	No. of records	Abbreviation
Balleny Islands region	1998.084	33	18.2	81	Ball
Fiji	1997.287	166	4.6	88	Fiji2
Santa Cruz Islands	1997.111	33	4.4	96	Sant
Flores	1996.169	587	7.3	90	Flor
Irian Jaya	1996.048	33	24.1	103	Iria
Sulawesi	1996.001	24	7.8	91	Sula
Mexico	1995.282	33	11.5	96	Mexi
Chile	1995.211	46	12.2	111	Chil
Kuril Islands	1994.277	54	30.0	101	Kuri
Bolivia	1994.160	631	26.3	88	Boli
Fiji	1994.068	562	3.1	72	Fiji
Macquarie Island	1989.143	10	13.6	52	Macq
Alaska	1987.334	10	7.3	41	Alas
Aleutians	1986.127	33	10.4	42	Aleu
Chile	1985.062	33	10.3	29	Chilo
New Ireland	1983.077	70	4.6	32	Nire
Banda Sea	1982.173	450	1.8	21	Band
New Hebrides	1980.199	33	4.8	21	Nheb
Colombia	1979.346	24	16.9	26	Colo
Kuril Islands	1978.340	91	6.4	24	Kurio
Sumbawa	1977.231	33	35.9	18	Sumb
Tonga	1977.173	65	13.9	16	Tong.

attention to those events which give a variance reduction greater than 70%. Most core-sensitive modes are also sensitive to mantle structure. Some events yield significantly different rotation angles for different mantle corrections^{29,30}. We regard this as an indication of serious noise contamination, and discard these events for a particular mode. This conservative choice sometimes results in very few events to constrain the rotation rate for individual modes; for example, only nine events constrain the rate for mode $3S_2$, and the rotation rate determined for this mode is less certain. We plot the apparent rotation angles for several modes as a function of event date (Fig. 3). The error bars represent the range of angles over which the fit to the receiver strips is adequate. The best-fitting straight line is also indicated, along with the corresponding apparent rotation rate.

For the hypothesis to be acceptable, all the modal splitting functions should appear to be rotating at the same rate. Figure 4 summarizes our best estimate of each mode's rotation rate. Also shown are the results of another recent mode study⁸ and body-wave studies^{1,3}. Even though some of the modes shown in Fig. 4 indicate significantly non-zero rotation rates (some indicate a slight westward inner-core rotation), these rates do not agree with each other. The mean rotation rate for all modes in our study is $0.01 \pm 0.21^\circ$ per year, which is obviously insignificantly different from zero. The results shown are obtained using model SB10L18³⁰ to perform the mantle corrections, but other models give rather similar results ($0.03 \pm 0.21^\circ$ per year for S16B30²⁹). Recent body-wave analyses suggest that the rotation rate may be as small as $0.2\text{--}0.3^\circ$ per year in an eastward sense. Such a rotation rate (also indicated in Fig. 4) is marginally consistent with most modes. We can certainly rule out a rate of 1° eastward, as well as a significant westward rotation. Perhaps complexity near the ray turning points or systematic changes in earthquake location procedures could allow smaller rotation rates to be compatible with the body-wave data.

We believe that a model analysis is the best way to determine inner-core rotation. First, we are dealing with large-scale vibrations which are insensitive to errors in event locations and to local structure in the inner core. Second, the method is independent of the earthquake source mechanism. Third, we do not need to worry about how much of the splitting functions we observe are caused by heterogeneity or anisotropy—all we care about is whether they change with time. Our results are marginally consistent with the small rotation rates reported in most recent body-wave analyses, though our best value indicates that the inner core is not rotating at a significant rate relative to the mantle. This would be in accord with the notion that the inner core is gravitationally locked to the mantle¹⁰. □

Received 11 May; accepted 6 August 1999.

1. Song, X. & Richards, P. G. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **382**, 221–224 (1996).
2. Su, W., Dziewonski, A. M. & Jeanloz, R. Planet within a planet: rotation of the inner core of the Earth. *Science* **274**, 1883–1887 (1996).
3. Creager, K. C. Inner core rotation rate from small-scale heterogeneity and time-varying travel times. *Science* **278**, 1248–1288 (1997).
4. Souriau, A. Earth's inner core—is the rotation real? *Science* **281**, 55–56 (1998).
5. Souriau, A. New seismological constraints on differential rotation rates of the inner core from Novaya Zemlya events recorded at DRV, Antarctica. *Geophys. J. Int.* **134**, F1–F5 (1998).
6. Souriau, A., Roudil, P. & Moynot, B. Inner core differential rotation: facts and artefacts. *Geophys. Res. Lett.* **24**, 2103–2106 (1997).
7. Creager, K. C. Large-scale variations in inner core anisotropy. *J. Geophys. Res.* (in the press).
8. Sharrock, D. S. & Woodhouse, J. H. Investigation of time dependent inner core structure by the analysis of free oscillation spectra. *Earth Planets Space* **50**, 1013–1018 (1998).
9. Masters, G., Laske, G. & Gilbert, F. Autoregressive estimation of the splitting matrix of free-oscillation multiplets. *J. Geophys. Res.* (submitted).
10. Buffet, B. A. A mechanism for decade fluctuations in the length of day. *Geophys. Res. Lett.* **23**, 3803–3806 (1996).
11. Masters, G. & Gilbert, F. Structure of the inner core inferred from observations of its spheroidal shear modes. *Geophys. Res. Lett.* **8**, 569–571 (1981).
12. Woodhouse, J. H., Giardini, D. & Li, X.-D. Evidence for inner core anisotropy from free oscillations. *Geophys. Res. Lett.* **13**, 1549–1552 (1986).
13. Morelli, A., Dziewonski, A. M. & Woodhouse, J. H. Anisotropy of the inner core inferred from PKIKP travel times. *Geophys. Res. Lett.* **13**, 1545–1548 (1986).
14. Poupinet, G., Pilet, R. & Souriau, A. Possible heterogeneity of the earth's core deduced from PKIKP travel times. *Nature* **305**, 204–206 (1983).

15. Jeanloz, R. & Wenk, H.-R. Convection and anisotropy of the inner core. *Geophys. Res. Lett.* **15**, 72–75 (1988).
16. Karato, S. Inner core anisotropy due to magnetic field-induced preferred orientation of iron. *Science* **262**, 1708–1711 (1993).
17. Yoshida, S., Sumita, I. & Kumazawa, M. Growth model of the inner core coupled with the outer core dynamics and the resulting elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103 (1997).
18. Bergman, M. I. Measurements of elastic anisotropy due to solidification texturing and the implications for the Earth's inner core. *Nature* **389**, 60–63 (1997).
19. Su, W.-J. & Dziewonski, A. M. Inner core anisotropy in three dimensions. *J. Geophys. Res.* **100**, 9831–9852 (1995).
20. Shearer, P. M. & Toy, K. M. PKP(BC) versus PKP(DF) differential travel times and aspherical structure in Earth's inner core. *J. Geophys. Res.* **96**, 2233–2247 (1991).
21. McSweeney, T. J., Creager, K. C. & Merrill, R. T. Depth extent of inner core seismic anisotropy and implications for geomagnetism. *Phys. Earth. Planet. Inter.* **101**, 131–156 (1997).
22. Song, X. Anisotropy of the Earth's inner core. *Rev. Geophys.* **35**, 297–313 (1997).
23. Woodhouse, J. H. The coupling and attenuation of nearly resonant multiplets in the earth's free oscillation spectrum. *Geophys. J. R. Astron. Soc.* **61**, 261–283 (1980).
24. Woodhouse, J. H. & Dahlen, F. A. The effect of a general aspherical perturbation on the free oscillations of the earth. *Geophys. J. R. Astron. Soc.* **53**, 335–354 (1978).
25. Woodhouse, J. H. & Giardini, D. Inversion for the splitting function of isolated low order normal mode multiplets. *Eos* **66**, 300 (1985).
26. Ritzwoller, M., Masters, G. & Gilbert, F. Observations of anomalous splitting and their interpretation in terms of aspherical structure. *J. Geophys. Res.* **91**, 10203–10228 (1986).
27. Giardini, D., Li, X.-D. & Woodhouse, J. H. Splitting functions of long period normal modes of the earth. *J. Geophys. Res.* **93**, 13716–13742 (1988).
28. Li, X.-D., Giardini, D. & Woodhouse, J. H. Large-scale three-dimensional even-degree structure of the Earth from splitting of long-period normal modes. *J. Geophys. Res.* **96**, 551–577 (1991).
29. Masters, G., Johnson, S., Laske, G. & Bolton, H. A shear velocity model of the mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1385–1411 (1996).
30. Masters, G., Bolton, H. & Laske, G. Joint seismic tomography for P and S velocities: how pervasive are chemical anomalies in the mantle? *Eos* **80**, S14 (1999).

Acknowledgements

The data used in this study were collected at a variety of global seismic networks and obtained from the IRIS-DMC, GEOSCOPE and BFO. This research was supported by the US NSF.

Correspondence and requests for materials should be addressed to G.L. (e-mail: glaske@ucsd.edu).

Environmental warming alters food-web structure and ecosystem function

Owen L. Petchey, P. Timon McPhearson, Timothy M. Casey & Peter J. Morin

Department of Ecology, Evolution and Natural Resources, 14 College Farm Road, Cook College, Rutgers University, New Brunswick, New Jersey 08901-8511, USA

We know little about how ecosystems of different complexity will respond to global warming^{1–5}. Microcosms permit experimental control over species composition and rates of environmental change. Here we show using microcosm experiments that extinction risk in warming environments depends on trophic position but remains unaffected by biodiversity. Warmed communities disproportionately lose top predators and herbivores, and become increasingly dominated by autotrophs and bacterivores. Changes in the relative distribution of organisms among trophically defined functional groups lead to differences in ecosystem function beyond those expected from temperature-dependent physiological rates. Diverse communities retain more species than depauperate ones, as predicted by the insurance hypothesis, which suggests that high biodiversity buffers against the effects of environmental variation because tolerant species are more likely to be found^{6,7}. Studies of single trophic levels clearly show that warming can affect the distribution and abundance of species^{2,4,5}, but complex responses generated in entire food webs greatly complicate inferences based on single functional groups. We used microcosms containing aquatic microbes to investigate

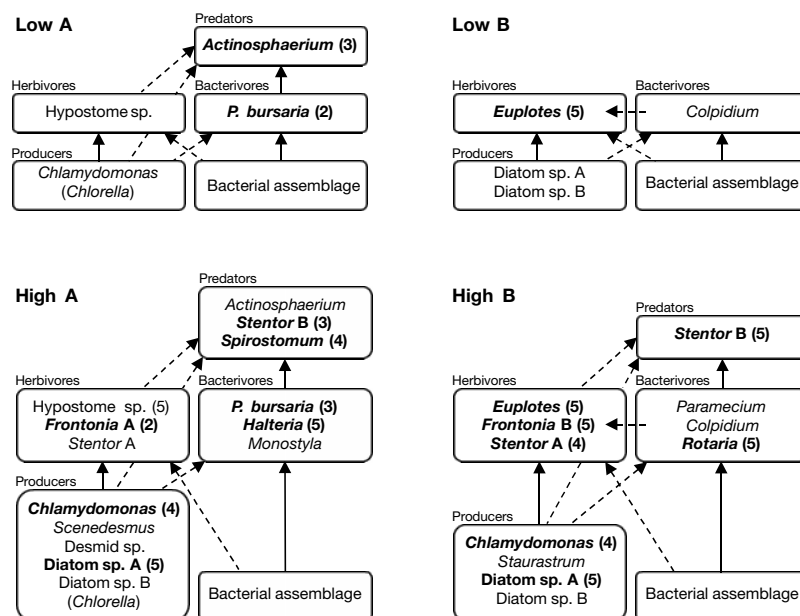


Figure 1 Species combinations and trophic structure used to manipulate diversity and composition. Bold type indicates species that were adversely affected by warming and numbers in parentheses indicate the number of replicates (out of five) in which the species became extinct. Three ubiquitous species are not shown: a cyanobacterium (producer); small *Amoeba* sp. (bacterivore); and microflagellates (bacterivore). *Paramecium bursaria*

how ecosystems of different complexity respond to simulated environmental warming. Different replicated communities experienced either constant or gradually increasing temperature regimes. Communities included several trophic groups (Fig. 1), because interactions within and among functional groups may modify responses to environmental change^{1–3,8}. Two different species compositions within both levels of species richness separated the effects of species richness from those of species identity. The experiment lasted seven weeks, corresponding to roughly 50–100 protist generations. Warming occurred at +2 °C per week, or about 0.1–0.2 °C per generation. Some global climate change models predict a +2 °C temperature increase over the next 100 years⁹, so our warming treatment scales reasonably with rates of temperature change that long-lived organisms might experience. Responses of short-lived microorganisms also reflect long-term dynamics rather than transient consequences of initial conditions^{6,10,11}.

Warmed communities lost more species (~30–40%) than communities in constant environments (~18%; Fig. 2a). Diversity and species composition did not significantly affect extinction frequencies. More-diverse communities were not especially fragile^{12,13} and did not exhibit the cascading extinctions predicted by models of self-organized systems^{14,15}. Higher biodiversity provided partial insurance against the complete loss of functional groups^{6,16}, basically because all communities lost a similar fraction of their initial species richness.

Warming greatly increased herbivore and predator extinction frequencies but had little effect on producers and bacterivores (Fig. 2b), leaving producers and bacterivores over-represented in warmed communities. Herbivores and predators became extinct more frequently than producers and bacterivores in constant-temperature communities (Fig. 2b).

Manipulations of initial diversity, species composition and environmental warming altered the biomass of several trophic groups measured during week six (diversity × composition × warming interaction; multivariate analysis of variation (MANOVA) $P = 0.039$). These patterns were consistent with cascading effects of altered trophic structure in which decreases in consumers cause increases in prey abundances^{17,18}, and mirror the results of

is listed as a bacterivore, but also contains photosynthetic endosymbionts (*Chlorella*), shown in the producer trophic group in parentheses. Solid arrows indicate definite trophic relations based on direct observations of feeding, and dashed arrows indicate tentative trophic relations inferred from the natural history of the organisms.

manipulations of heterotroph diversity in comparable systems^{6,19}. Warming increased producer and bacterivore biomass, while idiosyncratically affecting total biomass (Fig. 3). Warming did not alter bacterial biomass, which was consistent with the persistence of bacterivores in all treatments (Fig. 3). Greater total biomass in high-diversity communities (Fig. 3) could reflect sampling effects of biodiversity and/or overyielding resulting from complementary resource use^{20–23}.

Warming significantly increased primary production (Fig. 4a), directly through increased temperature-dependent physiological rates and indirectly through changes in trophic structure. Higher

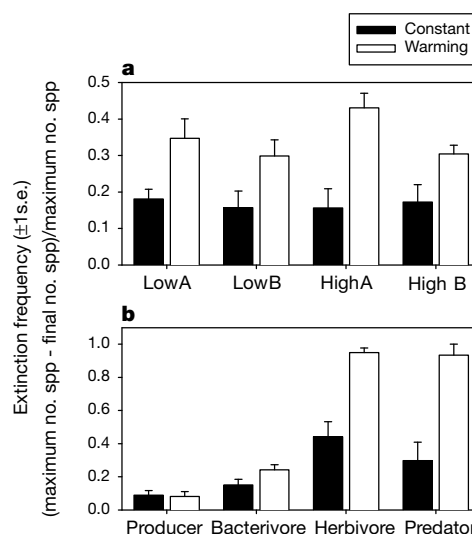


Figure 2 Effects of warming on extinctions. **a**, Treatment extinction frequencies. Warming significantly increases extinction frequency (ANOVA, $n = 40$, main effect of temperature, $P < 0.0001$), but no other main effects or interaction terms are significant. **b**, Trophic position extinction frequencies. Trophic position, temperature treatment and their interaction all significantly affect extinction frequency (ANOVA, $n = 149$, $P < 0.0001$).

primary production in less-diverse communities (Fig. 4a) was consistent with patterns seen in other aquatic studies^{6,19}. Altered trophic structure (summarized by principal component axis 1, Fig. 4b) explained deviations in observed production from values predicted by changing temperatures. Increased dominance of producers and bacterivores in warmed communities was correlated with elevated production¹⁹. Increased producer dominance should increase primary productivity, but the mechanisms linking increased bacterivore dominance to increased primary production are unclear. Warming affected primary production similarly in all four community types, even though diverse communities retained more species, perhaps because changes in trophic structure were qualitatively similar across communities. Warming also increased decomposition (Fig. 4c), probably through direct effects on decomposer physiology and indirect effects on food-web structure. Initial food-web structure affected decomposition (Fig. 4c), as previously found¹¹. To the extent that processes in our microcosms mirror processes operating in more complex natural systems, it seems likely that warming could significantly alter carbon flows in aquatic ecosystems by increasing decomposition. Although consistent across the four communities that we studied, warming might have different effects on food-web structure and ecosystem function in communities with other combinations of species.

Some theories predict that biodiversity will promote ecosystem integrity in changing climates, because high diversity ensures that functional groups will retain at least one species able to tolerate

altered conditions^{6,7,16}. More-diverse systems may also be less resistant to environmental change if population dynamics are less resilient in diverse systems^{12,13}, and especially if the extinction cascades predicted by models of self-organized systems occur^{14,15}. Our results do not indicate that greater diversity increased either fragility or resilience; however, diverse communities apparently contained more species able to survive in and respond to changed conditions. Seemingly redundant species may become important when ecosystems experience environmental stresses⁷.

Decreased heterotroph diversity releases top-down control of producers and increases primary production in studies ranging in scale from microcosms^{6,19} to whole lakes²⁴. Our results agree with other microcosm studies, despite differences in initial conditions and details of culture procedures^{6,19}. The greater extinction frequency for species in higher trophic positions is correlated with their larger body sizes (logistic regression, $\chi^2 = 25.8$, $P < 0.0001$) or lower abundances (logistic regression, $\chi^2 = 17.5$, $P < 0.0001$). Studies of the impact of global warming on natural systems also point to a preponderance of extinctions high in the food web²⁵,

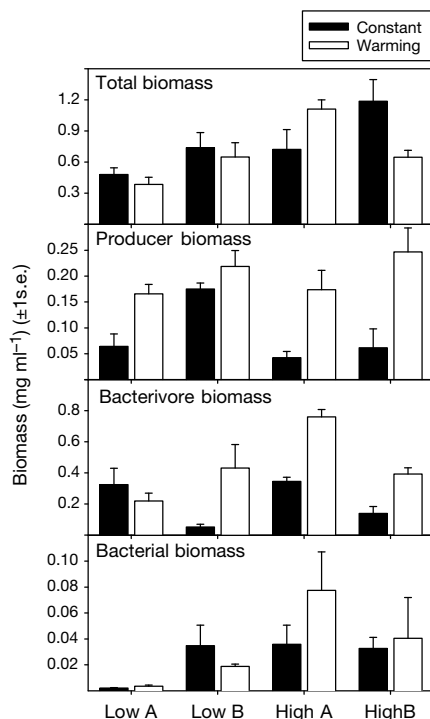


Figure 3 Biomass of different trophic groups during the sixth week. Total biomass was significantly higher in more-diverse communities (ANOVA, $n = 39$ (one outlier was excluded from the biomass analyses), main effect of diversity, $P < 0.001$; the composition $\times$ temperature and composition $\times$ temperature $\times$ diversity interaction terms were both significant at $P < 0.05$). Producer biomass was significantly higher in warmed communities (ANOVA, $n = 39$, main effect of warming, $P < 0.0001$) and in composition B communities ($P < 0.01$). Bacterivore biomass was significantly higher in warming communities (ANOVA, $n = 39$, main effect of warming, $P < 0.0001$). Diversity, composition, their interaction term and the three-way interaction term were also significant in ANOVA of bacterivore biomass. Bacterial biomass was not significantly different in the warming treatment (ANOVA, $n = 39$, main effect of warming, $P = 0.42$) but was significantly higher in more-diverse communities (ANOVA, $n = 39$, main effect of diversity, $P < 0.05$).

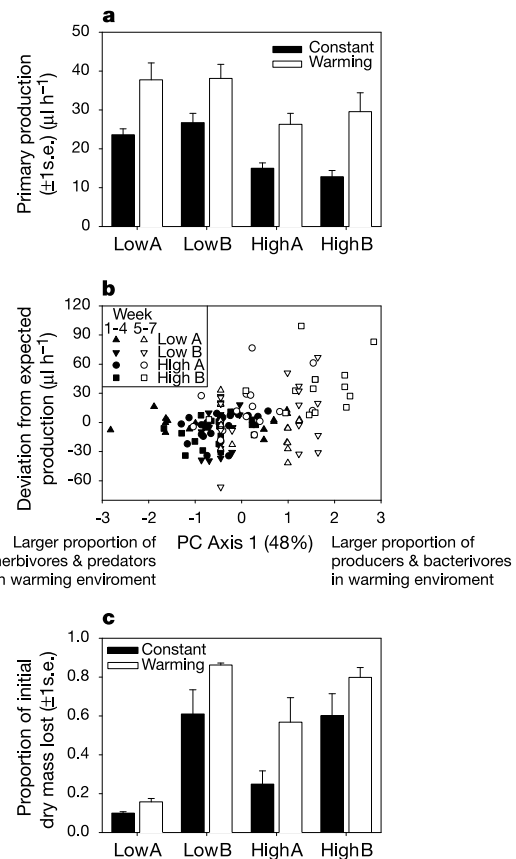


Figure 4 Impacts of warming on ecosystem properties. **a**, Mean primary production (mean across replicates of within replicate time average). Diversity (ANOVA, $n = 40$, $P < 0.0001$) and temperature regime (ANOVA, $n = 40$, $P < 0.0001$) significantly affected primary production. **b**, Altered trophic structure is closely related to changes in production in warmed communities. Text at the ends of the x axis shows the meaning of the principal component (PC) axis found from significant correlation between PC scores and the four original proportional variables. The percentage in parentheses indicates the amount of variance explained by PC axis 1. Only PC axis 1 (parameter estimate = 8.9 ± 2.3 , $P < 0.0001$) and PC axis 3 (parameter estimate = -4.9 ± 1.9 , $P < 0.05$) are significant in the multiple regression (note that the intercept is not significantly different from zero ($P = 0.51$) indicating that Q_{10} is not significantly different from 2). **c**, Decomposition after five weeks was higher in warming communities (ANOVA, $n = 40$, $P < 0.05$) and was also affected by composition (ANOVA, $n = 40$, $P < 0.001$), diversity (ANOVA, $n = 40$, $P < 0.05$) and the interaction between composition and diversity (ANOVA, $n = 40$, $P < 0.01$).

although the proximal mechanism for these losses remains uncertain. Beyond endangering top predators, warming caused complex changes in trophic structure that altered ecosystem processes, for example, facilitation of primary production resulting from reduced top-down control of producers. Comparable changes in trophic structure and ecosystem processes could shift the carbon budget of natural aquatic ecosystems²⁶, producing an important feedback between global warming and food-web structure. □

Methods

Microcosms

Microcosms were 250-ml culture bottles containing 100 ml of medium¹¹ inoculated with either 6–7 (low-diversity) or 13–16 (high-diversity) species of eukaryotic microorganisms. Different combinations of species (Fig. 1) created two types of low-diversity community (Low A and Low B) and two types of high-diversity community (High A and High B). A predator was added to the Low B community but failed to establish. Additions of bacteria (*Serratia marcescens*, *Bacillus subtilis* and *Bacillus cereus* in all microcosms) and other microorganisms were lagged to ensure that consumers were not introduced until their prey became abundant.

Directional environmental change

Half the microcosms experienced warming of +2 °C per week for the first five weeks of the experiment (22–32 °C), constant temperature for the sixth week, and warming to 34 °C at the beginning of the seventh week. The other half remained at 22 °C for the duration of the experiment. Replicates of temperature treatments were housed in four incubators (two constant temperature, two warming) with 16:8 h light:dark cycles to avoid pseudoreplication of temperature regimes.

Community monitoring

Species presence/absence was determined weekly by microscopy. We calculated extinction frequency as (maximum species richness – final species richness)/maximum species richness, because some species grew slowly to detectable densities. No contamination occurred. Population density of each species was estimated during week six by counting the number of individuals in a sample and converting counts to biomass using species-specific estimates of cell mass²⁷. Bacterial density was estimated weekly from plate counts.

Primary production

A multichannel closed-circuit respirometer (Columbus Scientific) measured O₂ flux over 4 h with lights on and 4 h with lights off each week. Microcosms were housed in heated and lit water baths during respirometry to maintain temperature treatments. Primary production was O₂ flux rate (light) – O₂ flux rate (dark). Expected values of primary production based on temperature-dependent physiological rates were estimated as $\text{Production}_{\text{expected}} = \text{Production}_{\text{constant}} \times Q_{10} \times (\text{Temperature}_{\text{warmed}} - \text{Temperature}_{\text{constant}})/10$ for each replicate during each week with $Q_{10} = 2$. Deviation of observed from expected production under warming was the difference between observed and expected values on a given date.

Decomposition

We estimated decomposition from the proportion change in dry mass over the first five weeks of preweighed wheat seeds in individually labelled nylon mesh bags¹.

Data analysis

Effects of warming, diversity and composition on extinction frequencies, primary production and decomposition were determined with separate three-way analyses of variance (ANOVAs). Effects of temperature regime and trophic position on extinction frequencies were determined with a two-way ANOVA. Effects of initial diversity, species composition and warming on the biomass of five trophic groups (bacteria, producers, bacterivores, herbivores and predators) were determined with MANOVA. The differences between warming and constant communities in proportions of producers, bacterivores, herbivores and predators quantified the effects of warming on trophic structure, and principal components analysis summarized variation in these four differenced proportion variables. Multiple regression related the deviation in production to changes in trophic structure (the first three PC axes), total number of species and bacterial biomass. Separate logistic regressions tested for relations between extinctions (binary variable) and log (body size), and extinctions and log (abundance) because body size and abundance were tightly correlated ($r^2 = 0.90$).

Received 19 July; accepted 7 September 1999.

1. Frost, T. M., Carpenter, S. R., Ives, A. R. & Kratz, T. K. In *Linking Species and Ecosystems* (eds Jones, C. G. & Lawton, J. H.) 224–239 (Chapman & Hall, New York, 1995).
2. Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B. & Wood, S. Making mistakes when predicting shifts in species range in response to global warming. *Nature* **391**, 783–786 (1998).
3. Sanford, E. Regulation of keystone predation by small changes in ocean temperature. *Science* **283**, 2095–2097 (1999).
4. Harte, J. & Shaw, R. Shifting dominance within a montane vegetation community: results of a climate-warming experiment. *Science* **267**, 876–880 (1995).
5. Parmesan, C. et al. Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).

6. Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
7. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
8. Ives, A. R. & Gilchrist, G. In *Biotic Interactions and Global Change* (eds Kareiva, P. M., Kingsolver, J. G. & Huey, R. B.) 120–146 (Sinauer Associates Inc., Sunderland, 1993).
9. Houghton, J. T. et al. *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, Cambridge, 1996).
10. Lawton, J. H. Ecological experiments with model systems. *Science* **269**, 328–331 (1995).
11. McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
12. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, Princeton, 1973).
13. Pimm, S. L. *The Balance of Nature?* (Univ. Chicago Press, Chicago, 1991).
14. Solé, R. V., Manrubia, S. C., Benton, M., Kauffman, S. & Bak, P. Criticality and scaling in evolutionary ecology. *Trends. Ecol. Evol.* **14**, 156–160 (1999).
15. Bak, P. *How Nature Works* (Springer, New York, 1996).
16. Naeem, S. Species redundancy and ecosystem reliability. *Cons. Biol.* **12**, 39–45 (1998).
17. Oksanen, L., Fretwell, S. D., Arruda, J. & Niemelä, P. Exploitation ecosystems in gradients of primary productivity. *Am. Nat.* **118**, 240–261 (1981).
18. Kaunzinger, C. M. K. & Morin, P. J. Productivity controls food-chain properties in microbial communities. *Nature* **395**, 495–497 (1998).
19. Naeem, S. & Li, S. Consumer species richness and autotrophic biomass. *Ecology* **79**, 2603–2615 (1998).
20. Hector, A. The effect of diversity on productivity: detecting the role of species complementarity. *Oikos* **82**, 597–599 (1998).
21. Loreau, M. Separating sampling and other effects in biodiversity experiments. *Oikos* **82**, 600–602 (1998).
22. Hooper, D. U. The role of complementarity and competition in ecosystem responses to variation in plant diversity. *Ecology* **79**, 704–719 (1998).
23. Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
24. Schindler, D. E., Carpenter, S. R., Cole, J. J., Kitchell, J. F. & Pace, M. L. Influence of food web structure on carbon exchange between lakes and the atmosphere. *Science* **277**, 248–251 (1997).
25. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
26. Del Giorgio, P. A., Cole, J. J., Caraco, N. F. & Peters, R. H. Linking planktonic biomass and metabolism to net gas fluxes in northern temperature lakes. *Ecology* **80**, 1422–1431 (1999).
27. McGrady-Steed, J. & Morin, P. J. Biodiversity, density compensation and the dynamics of populations and functional groups. *Ecology* (in the press).

Acknowledgements

We thank the NSF for supporting this research. Comments from the Morin lab, J. McGrady-Steed, S. Lawler, M. Loreau and S. Naeem improved the manuscript.

Correspondence and request for materials should be addressed to O.L.P. (e-mail: opetchey@rci.rutgers.edu).

A network of fast-spiking cells in the neocortex connected by electrical synapses

Mario Galarreta & Shaul Hestrin

Department of Anatomy and Neurobiology, University of Tennessee, Memphis, 855 Monroe Avenue, Memphis, Tennessee 38163, USA

Encoding of information in the cortex is thought to depend on synchronous firing of cortical neurons^{1,2}. Inhibitory neurons are known to be critical in the coordination of cortical activity^{3–5}, but how interaction among inhibitory cells promotes synchrony is not well understood^{4,6–12}. To address this issue directly, we have recorded simultaneously from pairs of fast-spiking (FS) cells, a type of γ -aminobutyric acid (GABA)-containing neocortical interneuron¹³. Here we report a high occurrence of electrical coupling among FS cells. Electrical synapses were not found among pyramidal neurons or between FS cells and other cortical cells. Some FS cells were interconnected by both electrical and GABAergic synapses. We show that communication through electrical synapses allows excitatory signalling among inhibitory cells and promotes their synchronous spiking. These results indicate that electrical synapses establish a network of fast-spiking cells in the neocortex which may play a key role in coordinating cortical activity.

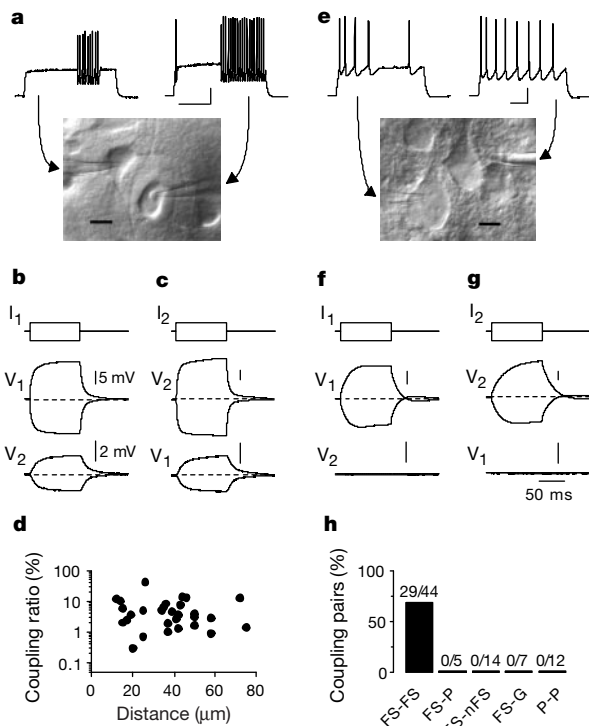


Figure 1 Specific electrical coupling among fast-spiking cells. **a**, Pattern of spiking and IR-DIC video image of two electrically coupled FS cells. **b**, **c**, Simultaneous depolarization or hyperpolarization of two FS cells (V_1 , V_2) when positive or negative currents were injected in only one of them (I_1 or I_2). **d**, Coupling ratio versus the distance between the somata of electrically coupled FS cells ($n = 29$). **e**, **f**, **g**, Pairs of neighbouring pyramidal neurons were not electrically coupled. **h**, FS cells were only electrically coupled to other FS cells. FS, fast-spiking cells; P, pyramidal cells; nFS, non-FS, non-pyramidal cells; G, glial cells. Scale: **a**, **e**, vertical 20 mV, horizontal 100 ms (spiking pattern), horizontal 10 μ m (IR-DIC image); **c**, **f**, **g** as in **b**.

We recorded simultaneously from pairs of fast-spiking non-pyramidal cells (FS cells) in brain slices (Fig. 1a). Fast-spiking cells in layer V were identified on the basis of their characteristic pattern of spiking in response to current injection^{14–17} (Fig. 1a; see Methods). Additionally, we found that FS cells showed, in most cases, parvalbumin immunoreactivity^{14,18} and had a local axonal projection and aspiny radial dendrites. After identifying a pair of FS cells we tested their electrical coupling under current-clamp or voltage-clamp conditions. Under current-clamp conditions, hyperpolarization of one of the neurons (V_1 in Fig. 1b) by injecting current into this cell (I_1) produced a simultaneous hyperpolarization of the non-injected neuron (V_2 in Fig. 1b). Similarly, depolarizing cell 1 produced a depolarization in cell 2. As expected for electrotonic propagation, voltage deflections recorded in the non-injected neuron had a smaller amplitude and slower time course than those in the injected cell. In all cases, electrical transmission was found to be reciprocal (Fig. 1c). We found that the frequency of coupling among FS cells was 66% (29 of 44 pairs tested, somata less than 80 μ m apart). The coupling ratio, estimated as the ratio of the amplitude of the voltage change in the non-injected cell to that in the injected cell, had a mean of $6.4 \pm 1.7\%$ (Fig. 1d; range: 0.3–41.1%, $n = 29$). In three further experiments with pairs of FS cells separated by 170–270 μ m, one pair was also found to be electrically coupled (coupling ratio: 1.4%).

Electrical coupling appears to be a cell-type specific feature. We found no evidence of electrical coupling among pairs of neighbouring pyramidal neurons (Fig. 1e–g; average distance: $21.4 \pm 1.5 \mu$ m, $n = 12$ cells). Also, we have not observed electrical coupling in experiments where dual recordings between FS cells and other types of cells were established ($n = 26$). These observations (Fig. 1h)

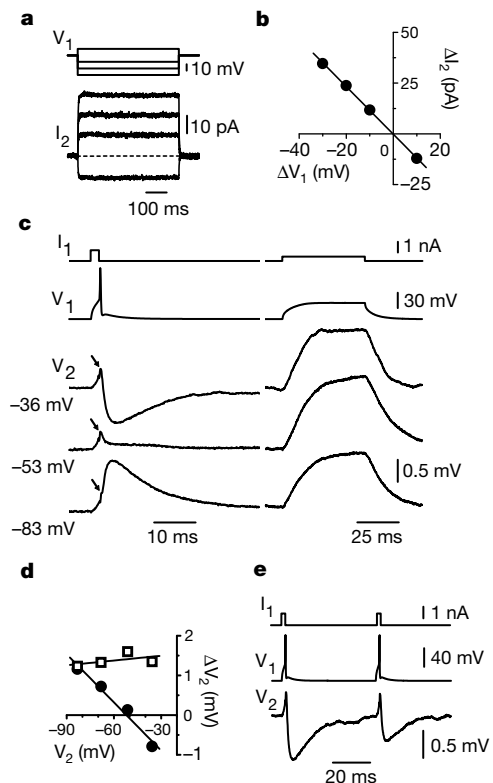


Figure 2 GABA-mediated synaptic transmission and electrical coupling among FS cells. **a**, **b** Example of the current (I_2) recorded from an FS cell when stepping the voltage (V_1) of another FS cell. The estimated conductance of this electrical connection was 1.17 nS. **c**, **d**, Different voltage-dependence of GABA-mediated uIPSPs (filled circles) and electrical transmission (open squares) in a pair of FS cells connected by both electrical and chemical synapses. **e**, Paired-pulse depression of the uIPSPs between the same pair of FS cells as in **c**.

included simultaneous recordings of FS cells and other types of nonpyramidal cells (non-FS nonpyramidal cells, $n = 14$), FS cells and pyramidal cells ($n = 5$), and FS cells and glial cells ($n = 7$).

To estimate the conductance between electrically coupled FS cells we used voltage-clamp conditions¹⁹. The membrane potential of both FS cells was initially held at -60 mV. We then measured the current produced in cell 2 following a range of voltage commands at cell 1 (Fig. 2a). The conductance between coupled cells was non-rectifying at the membrane potentials tested (Fig. 2b) and its mean amplitude was 658 ± 175 pS ($n = 20$; range: 30–3364 pS). The time course of the current recorded in the coupled neuron was fitted with an exponential function whose time constant ranged from 1.2 ms to 18.5 ms (mean: 6.2 ± 1.4 ms, $n = 20$ pairs). This wide range suggests that electrical contacts occur at different electrotonic distances. In addition, in some pairs we observed that the coupling time constant was significantly faster than the effective input time constant of both cells ($n = 11$ pairs), indicating that in these pairs electrical contacts may occur at proximal locations.

Eight of 44 (18%) pairs of FS cells were connected via GABA-mediated synaptic transmission. Two of these pairs were reciprocally connected and all the pairs connected by GABAergic synapses were also electrically coupled. The time course of the unitary inhibitory postsynaptic potential (uIPSP) had an exponential decay time constant of 10.4 ± 2.0 ms ($n = 7$) and exhibited paired-pulse depression¹⁰ at an interstimulus interval of 50 ms (uIPSP2/uIPSP1 = 0.74 ± 0.03 , $n = 7$; Fig. 2e). The average conductance produced by the inhibitory postsynaptic currents (IPSCs) was 1.7 ± 0.4 nS ($n = 4$). When both chemical and electrical transmission were present between two FS cells, a presynaptic spike generated a dual component response in the postsynaptic

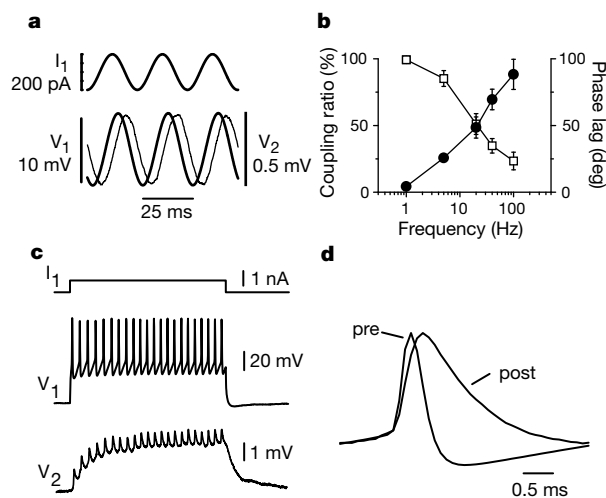


Figure 3 Electrical coupling behaving as a low-pass filter. **a**, Phase lag (62°) between the membrane voltage oscillations produced in a pair of electrically coupled FS cells (V_1 and V_2) when injecting a sine wave current (40 Hz) in one of them (I_1). **b**, Frequency dependence of sine wave transmitted through electrical synapses ($n = 3-6$). Coupling ratios are percentages of the DC signals. Squares, coupling ratio; circles, phase lag. **c**, Example of the electrotonic transmission of a train of action potentials between two coupled FS cells. **d**, Scaled superimposition of a presynaptic spike ('pre') and the corresponding response in the coupled cell ('post').

cell (Fig. 2c). The electrotonic signal consisted of an initial slow response reflecting the subthreshold depolarization of the 'presynaptic' cell, followed by a brief transient signal reflecting the presynaptic spike (Fig. 2c, arrows). When the membrane potential of the postsynaptic cell was maintained at -35 to -40 mV, the GABA_A-receptor mediated IPSP was hyperpolarizing whereas the electrotonic component was depolarizing (Fig. 2c and d). The peak of the electrotonic depolarization occurred earlier than that of the GABA_A-receptor mediated IPSP. Near the reversal potential of the IPSP only a depolarizing electrotonic component was apparent (Fig. 2c and d). At a more hyperpolarized potential both chemical and electrotonic signals were depolarizing (Fig. 2c and d).

We found that the efficacy of signal transmission through electrical coupling is frequency-dependent. We injected sinusoidal current waveforms into an FS cell and measured the response in the coupled cell (Fig. 3a). Responses to low-frequency sine waves showed larger coupling ratios and smaller phase lags than those induced by high-frequency sine waves (Fig. 3b). This frequency-dependent attenuation resulted in a smaller coupling ratio of action potentials compared with the coupling ratio of DC signals (1.4% and 10%, respectively; $n = 10$). When a train of spikes was generated in an FS cell (Fig. 3c), the response in the coupled cell consisted of a slow and relatively large response to the DC component of the spike train and smaller transient responses corresponding to each action potential (Fig. 3c). The average latency between the peak of a presynaptic spike (pre) and the peak of its corresponding response in the coupled cell (post) was 0.34 ± 0.02 ms ($n = 10$, Fig. 3d).

Next, we investigated whether electrical coupling among individual FS cells could facilitate simultaneous spiking. To address this issue, a close-to-threshold depolarizing current pulse was injected at different times in two electrically coupled FS cells (Fig. 4a, open triangles 1 and 2). In most cases, these asynchronous pulses did not generate an action potential in either neuron. However, when the same current pulses were synchronously injected in both cells, the two neurons reached spike threshold (Fig. 4a, filled triangle 1 + 2; Fig. 4b). Similar observations were obtained in four pairs of electrically coupled FS cells whose coupling ratios ranged from 4.4% to 41.0%. These results indicate that electrical coupling can

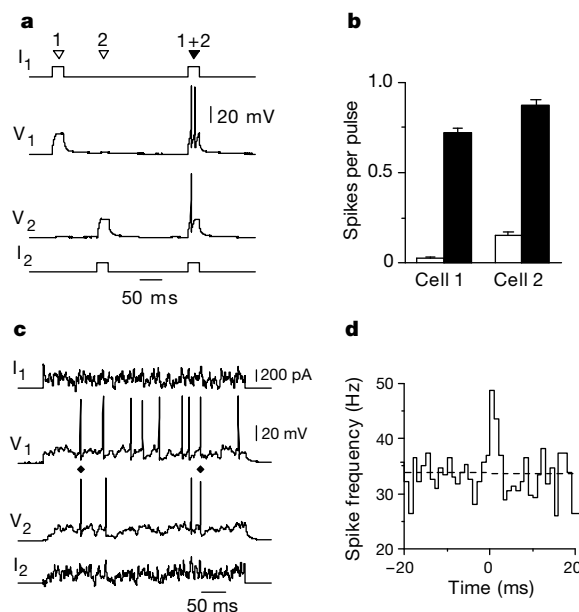


Figure 4 Electrical coupling promoting simultaneous spiking. **a**, Synchronous injection (filled triangle) of subthreshold current pulses (open triangles) in two electrically coupled FS cells increases the likelihood of firing. Traces represent single trials. **b**, Averaged data ($n = 264$ trials) from the same experiment as in **a**. White bars, non synchronized pulses; black bars, synchronized pulses. **c**, Two electrically coupled FS-cells were injected with uncorrelated synthetic random currents. Diamonds indicate spikes occurring in both cells within a 1-ms window. Traces are single trials. **d**, Normalized cross-correlogram analysis of the pair of cells shown in **c** ($n = 918$ trials). Bin size is 1 ms. Spike frequency is increased near 0 ms.

promote action potential generation. We also wanted to determine whether electrical coupling could facilitate synchronous spiking in response to a 'natural' stimulus. *In vivo* the activity of cortical neurons is ongoing²⁰; therefore, neurons are bombarded with both excitatory and inhibitory postsynaptic currents (EPSCs and IPSCs) at high frequency. We simulated natural activity by injecting two electrically coupled cells with uncorrelated random signals obtained by convolving Poisson trains and previously recorded EPSC and IPSC waveforms (Fig. 4a; see Methods). A cross-correlogram of action potentials evoked in these cells showed a significant ($P < 0.05$) increase in firing frequency centred close to 0 ms (Fig. 4d; $n = 4$ pairs). Rapid electrical transmission of action potentials among individual cells can thus produce precise (within 1 ms) coherent spiking in response to uncorrelated excitation.

Our data indicate that electrical synapses establish a functional network of specific inhibitory neurons within the neocortex. Previously, dye coupling was observed among neocortical interneurons^{9,21,22}, and anatomical studies observed gap junctions between nonpyramidal cells^{23,24}. The frequency dependence of electrical coupling suggests that effective transmission of action potentials will be limited to relatively small domains, whereas slow subthreshold waveforms could spread through larger numbers of electrically coupled FS cells. However, active propagation of spikes to the site of the electrical contact could result in a more efficient transmission than that predicted for passive high-frequency signals. *In vivo* and *in vitro* experiments^{4,12,25,26} have suggested that networks of inhibitory neurons are crucial in synchronizing neuronal activity. This proposal has been supported by theoretical models^{4,7,27-29}; however, synchronization depended on specific constraints²⁷⁻²⁹ when interneurons are interconnected by GABAergic synapses only. We propose that the existence of electrical coupling among inhibitory neurons, demonstrated here, may provide an additional mechanism that will facilitate precisely synchronous spiking of these cells and contribute to the control of activity in the cortex. □

Methods

Parasagittal cortical slices (300 μm thick, 30 degree angle) were obtained from Wistar rats (14–18 days old). Within this range, no correlation was found between the likelihood of coupling among FS cells and the animal age. The extracellular solution bathing the slices was kept at 32–33 °C and contained 125 mM NaCl, 2.5 mM KCl, 1.25 mM NaH_2PO_4 , 1 mM MgSO_4 , 2 mM CaCl_2 , 26 mM NaHCO_3 , 20 mM glucose, pH 7.4 (315 mOsm), and was continuously bubbled with a gas mixture of 95% O_2 and 5% CO_2 .

Cell identification

Cells were selected and identified based on their physiology as well as their immunocytochemistry and morphology. Cells in layer V of the somatosensory and the visual cortices were initially selected according to their morphology using infrared differential interference contrast (IR-DIC) optics. Neurons lacking an apical dendrite with multipolar appearance were selected as FS cell candidates. To distinguish FS cells from other types of non-FS nonpyramidal cell we examined their response to near-threshold current injection. The distinguishing feature that we used to classify FS cells was the generation of high-frequency non-accommodating discharges of action potentials in response to near-threshold current injection^{14–17}. We found that in FS cells ($n = 52$ cells), the mean interspike interval (ISI) in these near-threshold discharges was 8.0 ± 0.7 ms (range 2–23.7 ms), the coefficient of variation of the ISI was 0.085 ± 0.007 , and the ratio of the second ISI to the first was 1.08 ± 0.02 . Nonpyramidal neurons that lacked these characteristic discharges were classified as non-FS nonpyramidal cells. In this group, near-threshold discharges containing ≥ 3 spikes had a mean ISI of 62.3 ± 10.8 ms (range 31.6–161.8), a coefficient of variation of the ISI of 0.483 ± 0.096 , and a ratio of the second ISI to the first of 3.21 ± 1.00 ($n = 13$). FS cells also had the following characteristics: input resistance 92.4 ± 4 M Ω ($n = 52$), input time constant 8 ± 0.7 ms (measured by fitting the late phase of the response to a small current injection, $n = 50$), fast afterhyperpolarization amplitude -21.7 ± 0.6 mV (measured from the spike onset, $n = 52$), and spike width at threshold 0.6 ± 0.02 ms ($n = 52$). In addition, FS cells exhibited a remarkably high activity of spontaneous excitatory postsynaptic potentials (EPSPs). Pyramidal neurons were recognized by their characteristic morphology and pattern of spiking¹⁶. Glial cells were identified as having a resting potential more hyperpolarized than -80 mV, low input resistance, absence of synaptic inputs and inability to generate action potentials.

Morphology of neurons stained with biocytin was revealed by standard avidin-biotinylated-horseradish peroxidase complex (ABC, Vector Laboratories) and 3-3' diaminobenzidine reaction. Fast-spiking cells successfully recovered were nonpyramidal neurons with aspiny radial dendrites. The axon originated in most cases from the upper part of the soma and branched locally with a predominantly horizontal orientation. Immunofluorescence methods were used to detect the presence of parvalbumin in biocytin-stained cells. We used mouse monoclonal antibodies against parvalbumin (Swant, #235, 1/5000), and secondary mouse IgG antibodies (Fab specific) conjugated with TRITC (tetramethylrhodamineisothiocyanate) (Sigma, T-7782, 1/400). Biocytin-filled cells were detected using Neutravidin-Alexa 488 (Molecular Probes). Nine of twelve (75%) neurons classified as FS cells were immunopositive for parvalbumin. In contrast, none of ten non-FS cells tested for parvalbumin were immunopositive.

Recording and data analysis

Simultaneous somatic whole-cell recordings (Axopatch 200 A/B; Axon Instruments) were made using patch pipettes (2–4 M Ω) filled with a solution containing 130 mM K-methylsulphate, 6.3 mM KCl, 10 mM HEPES, 4 mM MgATP, 20 mM phosphocreatine(Na), 0.3 mM NaGTP, 10 mM EGTA and 0.3% mM biocytin, (pH 7.3, 295 mOsm). The liquid junction potential error was not corrected. The error in the estimation of the conductance between coupled cells produced by the series resistances was corrected¹⁹. Recordings from pairs of FS cells were obtained in the presence of DNQX (10 μM , RBI) to block AMPA/kainate receptor-mediated synaptic currents. Chemical synaptic transmission was studied by generating action potentials in the presynaptic neuron using brief pulses (3–5 ms) of suprathreshold currents. The reversal potential of the uIPSP was -56 mV ($n = 4$) and 10 μM bicuculline methiodide (a GABA_A-receptor antagonist) blocked the uIPSP generated by FS cells ($n = 3$). Voltage and current signals were filtered at 10 kHz and digitized at 16-bit resolution (National Instruments). The sampling frequency was 5–10 kHz. To synthesize 'natural' stimuli we convolved a Poisson train (2,000 Hz) with a waveform of an excitatory postsynaptic current (EPSC). The result was combined with the convolution of a Poisson train (1,000 Hz) and an IPSC waveform³⁰. Trials were separated by 1 s. For each trial a different randomly generated Poisson train was used. The waveforms of the EPSC and IPSC used in the convolution were previously recorded unitary currents.

Data are given as mean $\pm$ s.e.m. Statistical analysis testing two-sample hypothesis was performed using unpaired, two-tailed Student's t tests. Differences were considered statistically significant if $P \leq 0.05$. Peaks in the cross-correlogram were considered significant if individual bins exceeded expected value by 2.5 standard deviations.

Received 27 April; accepted 23 August 1999.

1. Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
2. Ritz, R. & Sejnowski, T. J. Synchronous oscillatory activity in sensory systems: new vistas on mechanisms. *Curr. Opin. Neurobiol.* **7**, 536–546 (1997).
3. Lytton, W. W. & Sejnowski, T. J. Simulations of cortical pyramidal neurons synchronized by inhibitory interneurons. *J. Neurophysiol.* **66**, 1059–1079 (1991).
4. Whittington, M. A., Traub, R. D. & Jefferys, J. G. R. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373**, 612–615 (1995).
5. Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378**, 75–78 (1995).

6. Michelson, H. B. & Wong, R. K. S. Synchronization of inhibitory neurones in the guinea-pig hippocampus in vitro. *J. Physiol. (Lond.)* **477**, 35–45 (1994).
7. Jefferys, J. G. R., Traub, R. D. & Whittington, M. A. Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208 (1996).
8. Strata, F. et al. A pacemaker current in dye-coupled hilar interneurons contributes to the generation of giant GABAergic potentials in developing hippocampus. *J. Neurosci.* **17**, 1435–1446 (1997).
9. Benardo, L. S. Recruitment of GABAergic inhibition and synchronization of inhibitory interneurons in rat neocortex. *J. Neurophysiol.* **77**, 3134–3144 (1997).
10. Tamás, G., Somogyi, P. & Buhl, E. H. Differentially interconnected networks of GABAergic interneurons in the visual cortex of the cat. *J. Neurosci.* **18**, 4255–4270 (1998).
11. Rinzel, J., Terman, D., Wang, X.-J. & Ermentrout, B. Propagating activity patterns in large-scale inhibitory neuronal networks. *Science* **279**, 1351–1355 (1998).
12. Zhang, Y. et al. Slow oscillations (≤ 1 Hz) mediated by GABAergic interneuronal networks in rat hippocampus. *J. Neurosci.* **18**, 9256–9268 (1998).
13. Kawaguchi, Y. & Kubota, Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb. Cortex* **7**, 476–486 (1997).
14. Kawaguchi, Y. & Kubota, Y. Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28k-immunoreactive neurons in layer V of rat frontal cortex. *J. Neurophysiol.* **70**, 387–396 (1993).
15. Kawaguchi, Y. Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex. *J. Neurosci.* **15**, 2638–2655 (1995).
16. Galarreta, M. & Hestrin, S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neurosci.* **1**, 587–594 (1998).
17. Angulo, M. C., Staiger, J. F., Rosier, J. & Audinat, E. Developmental synaptic changes increase the range of integrative capabilities of an identified excitatory neocortical connection. *J. Neurosci.* **19**, 1566–1576 (1999).
18. Cauli, B. et al. Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.* **17**, 3894–3906 (1997).
19. Neyton, J. & Trautmann, A. Single-channel currents of an intercellular junction. *Nature* **317**, 331–335 (1985).
20. Hubel, D. H. Single unit activity in striate cortex of unrestrained cats. *J. Physiol.* **147**, 226–238 (1959).
21. Connors, B. W., Benardo, L. S. & Prince, D. A. Coupling between neurons of the developing rat neocortex. *J. Neurosci.* **3**, 773–782 (1983).
22. Peinado, A., Yuste, R. & Katz, L. C. Extensive dye coupling between rat neocortical neurons during the period of circuit formation. *Neuron* **10**, 103–114 (1993).
23. Sloper, J. J. Gap junctions between dendrites in the primate neocortex. *Brain Res.* **44**, 641–646 (1972).
24. Katsumaru, H., Kosaka, T., Heizmann, C. W. & Hama, K. Gap junctions on GABAergic neurons containing the calcium-binding protein parvalbumin in the rat hippocampus (CA1 region). *Exp. Brain Res.* **72**, 363–370 (1988).
25. Bragin, A. et al. Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* **15**, 47–60 (1995).
26. Buhl, E. H., Tamás, G. & Fisahn, A. Cholinergic activation and tonic excitation induce persistent gamma oscillations in mouse somatosensory cortex in vitro. *J. Physiol. (Lond.)* **513**, 117–126 (1998).
27. Wang, X.-J. & Buzsáki, G. Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *J. Neurosci.* **16**, 6402–6413 (1996).
28. Traub, R. D., Whittington, M. A., Stanford, I. M. & Jefferys, J. G. R. A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* **383**, 621–624 (1996).
29. Bush, P. & Sejnowski, T. Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic networks models. *J. Comput. Neurosci.* **3**, 91–110 (1996).
30. Stevens, C. F. & Zador, A. M. Input synchrony and the irregular firing of cortical neurons. *Nature Neurosci.* **1**, 210–217 (1998).

Acknowledgements

We thank M. Chang for technical assistance. S.H. was supported by an NIH grant.

Correspondence and requests for materials should be addressed to S.H. (e-mail: shaul@nb.utm.edu).

Two networks of electrically coupled inhibitory neurons in neocortex

Jay R. Gibson*, Michael Beierlein* & Barry W. Connors

Department of Neuroscience, Division of Biology & Medicine, Brown University, Providence, Rhode Island 02912, USA

* These authors contributed equally to this work

Inhibitory interneurons are critical to sensory transformations, plasticity and synchronous activity in the neocortex^{1,2}. There are many types of inhibitory neurons, but their synaptic organization is poorly understood. Here we describe two functionally distinct inhibitory networks comprising either fast-spiking (FS) or low-threshold spiking (LTS) neurons. Paired-cell recordings showed that inhibitory neurons of the same type were strongly interconnected by electrical synapses, but electrical synapses between

different inhibitory cell types were rare. The electrical synapses were strong enough to synchronize spikes in coupled interneurons. Inhibitory chemical synapses were also common between FS cells, and between FS and LTS cells, but LTS cells rarely inhibited one another. Thalamocortical synapses, which convey sensory information to the cortex, specifically and strongly excited only the FS cell network. The electrical and chemical synaptic connections of different types of inhibitory neurons are specific, and may allow each inhibitory network to function independently.

We recorded data from two common types of inhibitory interneurons of somatosensory neocortex, and tested the possibility that each type participates in a functionally distinct network defined by its synaptic inputs, intrinsic membrane properties, chemical and electrical synaptic interconnections, and output targets. Recordings were performed in layers 4 and 6 of rat barrel cortex using an *in vitro* slice that preserved connections between thalamus and barrel cortex³ (Fig. 1a). We targeted neurons with large, usually vertically oriented somata (20–30 μm along the vertical axis, Fig. 1b) because these were most likely to be inhibitory⁴. Intrinsic membrane properties and synaptic inputs distinguished the two interneuron types^{5–7}. Fast-spiking cells had narrow action potentials (a mean of 0.35 ms at half-amplitude), each with a deep, brief after-hyperpolarization, high firing rates (up to 320 Hz) when stimulated, and little or no frequency adaptation (Fig. 1c, mean rates after 600 ms were 97% of initial rates). When stimulated near threshold, FS cells often generated a slowly depolarizing voltage ramp before firing spikes at long latencies (up to 600 ms); stronger stimuli evoked

spikes at short latencies. Low-threshold spiking cells had broader spikes (0.55 ms at half-amplitude), pronounced adaptation of firing frequency (mean rates after 600 ms fell to 39% of initial rates) and never generated a long delay before spiking (Fig. 1c). These LTS cells also displayed low-threshold spikes when depolarized from more negative potentials⁶. The two cell types had similar input resistances (FS, $58 \pm 26 \text{ M}\Omega$, $n = 127$; LTS, $71 \pm 37 \text{ M}\Omega$, $n = 40$; all data are expressed as mean $\pm$ s.d.).

Some interneurons in layer 4 were stained with biocytin or neurobiotin (FS, $n = 35$; LTS, $n = 5$). All had smooth or sparsely spiny dendrites, usually vertically oriented and bitufted, or multipolar, with dense axonal processes extending into layers 2/3 and 5. The morphologies of FS and LTS cells were not obviously different. Calcium-binding proteins and neuroactive peptides help to distinguish subtypes of γ -amino butyric acid (GABA)-mediated neurons

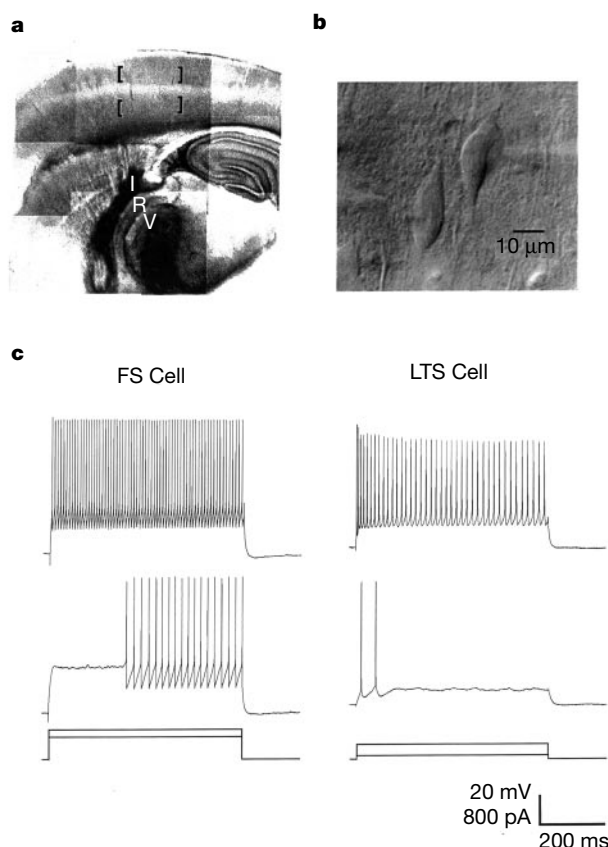


Figure 1 Two types of inhibitory interneurons. **a**, The thalamocortical slice as it appears in the recording chamber. Brackets mark the recording areas in layers 4 and 6 of neocortex. Thalamocortical afferents activated by stimulating the ventral basal nucleus of the thalamus (V) pass through the reticular nucleus of the thalamus (R) and the internal capsule (I). **b**, Video IR-DIC image of a pair of low-threshold spiking (LTS) cells just before recording. **c**, Current-clamp recordings from a fast-spiking (FS) cell (left) and an LTS cell (right) during injection of depolarizing current pulses of two intensities.

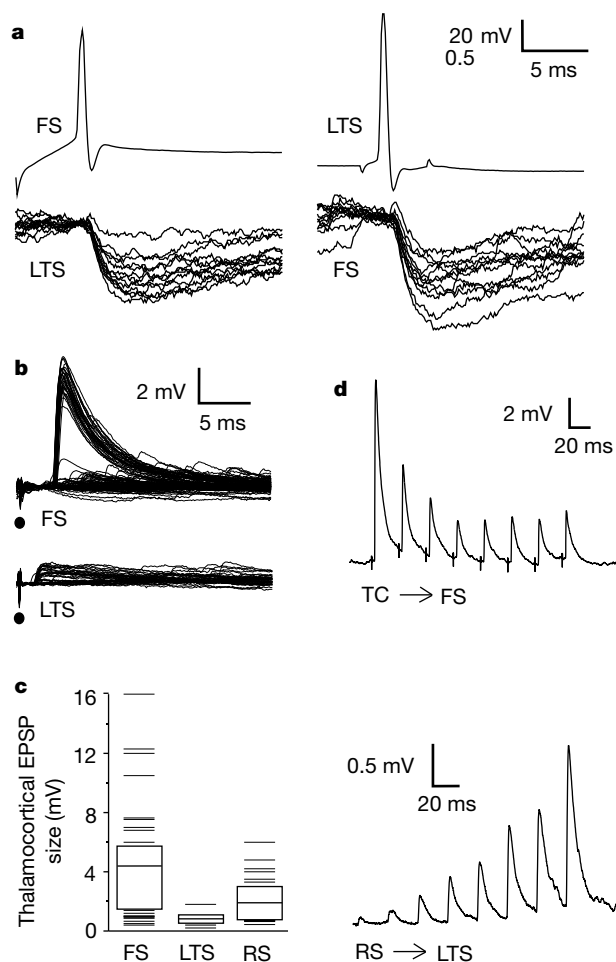


Figure 2 Chemical synaptic connections of inhibitory interneurons. **a**, Inhibitory synapses interconnecting FS and LTS neurons. During paired recordings, spikes in an FS cell evoked inhibitory postsynaptic potentials (IPSPs) in a neighbouring LTS cell (left); in a different pair, spikes in an LTS cell evoked IPSPs in an FS cell (right). In each case, 12 trials are illustrated from the postsynaptic cell, each aligned on the presynaptic spike that evoked it, and a single presynaptic spike is shown for each pair. **b**, Thalamocortical EPSPs evoked in an FS and an LTS cell with threshold stimuli. Multiple trials are overlaid. **c**, Summary of mean single-axon thalamocortical EPSP amplitudes, showing that FS cells responded more strongly than both LTS and regular-spiking (RS) cells. Boxes delineate the 25th and 75th percentiles, the line within the box is the sample mean, and lines outside the boxes plot the rest of the range ($n = 38$ FS, 8 LTS, 26 RS cells). **d**, Upper trace, thalamocortical (TC) inputs to FS cells strongly depress (average of 20 trials of suprathreshold stimuli applied at 40 Hz). Lower trace, intracortical excitatory inputs to LTS cells strongly facilitate (LTS cell response from presynaptic RS cell spikes evoked at 40 Hz; average of 20 trials).

Table 1 Electrical and chemical synapses between different types of neurons

Cell types	Total cell pairs	Chemical synapse	Reciprocal chemical synapses	Electrical synapse	Electrical and chemical synapses
FS → FS	39	24	6	24	17
LTS → LTS	13	2	0	11	1
LTS → FS	32	17	7	3	0
FS → LTS	32	11	7	3	0
RS → RS	80	9	0	0	0
FS/LTS → RS	71*	37	5	0	0
RS → FS/LTS	54*	22	5	0	0

Each row shows data from a particular type of neuronal pairing, with the number of cases that tested positively for a chemical synapse in one direction, a chemical synapse in two directions, electrical coupling, or both electrical and chemical synapses in at least one direction. FS, Fast-spiking cell (GABAergic); LTS, Low-threshold spiking cell (GABAergic); RS, Regular-spiking cell (glutamatergic). FS/LTS indicates that data from the two types were pooled. Arrows indicate the direction of the chemical synapses sampled.

*These represent the same sample of cell pairs, but in some cases glutamate receptors were blocked with AP5 and DNQX (see Methods), and thus the RS → FS/LTS connection could not be tested.

in neocortex². When immunocytochemistry was combined with intracellular biocytin staining, 6 out of 14 FS cells were parvalbumin immunoreactive, and 3 out of 6 LTS cells were somatostatin immunoreactive. No LTS cells were positive for parvalbumin ($n = 4$), and no FS cells were positive for somatostatin ($n = 8$). This labelling specificity is consistent with previous electrophysiological studies of these cell types^{6,7}.

Paired recordings confirmed that these interneurons were inhibitory. Spikes in presynaptic FS and LTS cells evoked inhibitory postsynaptic potentials (IPSPs) in postsynaptic neurons of all types (Fig. 2a; Table 1); however, inhibitory interconnections were much more prevalent between pairs of FS cells (62%) than between pairs of LTS cells (15%). Both FS cells and LTS cells also inhibited regular-spiking (RS) cells, which are excitatory neurons⁸. Regular-spiking neurons occasionally made excitatory chemical synapses with each other (11%), but RS cells more frequently excited the interneurons (41%).

Excitatory inputs to the two interneuron networks differed in several ways. Thalamocortical axons convey sensory information to the neocortex by excitatory synapses in layers 4 and 5/6 (ref. 1). They also evoke rapid and powerful feedforward inhibition, and physiological^{9,10} and anatomical¹¹ evidence suggests that FS cells may be involved. Single thalamocortical axons were stimulated, and excitatory postsynaptic potentials (EPSPs) were recorded (Fig. 2b). FS cells had, on average, the largest thalamocortical EPSPs (4.3 ± 3.7 mV, $n = 38$), whereas LTS cells had the smallest (0.8 ± 0.5 mV, $n = 8$; FS versus LTS, $p < 0.001$, t -test) (Fig. 2c). Thalamocortical responses were impossible to evoke in most LTS cells, even when neighbouring FS cells showed strong responses. The results indicate that LTS cells may receive weaker and far fewer thalamocortical inputs than FS cells. A less likely possibility is that afferents were differentially separated during slicing. Regular-spiking cells, the excitatory neurons, have thalamocortical responses of intermediate size (1.9 ± 1.5 mV, $n = 26$) (ref. 12). Thus, FS cells are very likely to be mediators of fast, feedforward thalamocortical inhibition.

Repetitive stimulation of thalamic axons at frequencies of 1–40 Hz produced strong depression of EPSPs in FS cells (Fig. 2d; mean EPSP amplitude fell by 73% after 8 stimuli at 10 Hz). Thalamocortical inputs to RS cells also depressed¹², but not as strongly as those to FS cells. Thalamic responses of LTS cells also clearly depressed, but they were too weak to measure. Both types of inhibitory cells received excitatory inputs from local RS cells (Table 1), and each had distinctive dynamics⁷. The RS–FS synapses depressed at all frequencies (mean EPSP amplitude fell by 46% after 8 stimuli at 10 Hz), whereas the synapses from RS to LTS cells markedly facilitated at frequencies above about 30 Hz (Fig. 2d; mean EPSP amplitude rose by 900% after 8 stimuli at 40 Hz).

The presence of dye coupling^{13–15} and gap junctions^{16,17} suggests

that neocortical interneurons may be electrically coupled. To assess the electrical strength, specificity and functional consequences of coupling, we made paired recordings from neighbouring interneurons ($<50 \mu\text{m}$ between somata). Among FS cell pairs, 62% were coupled (Table 1). Current steps applied to Cell 1 evoked voltage changes in Cell 2, and vice versa (Fig. 3a). The average attenuation of low-frequency voltages, or “coupling coefficient”¹⁸, between FS cells was 0.07 ± 0.06 ($n = 24$). Action potentials attenuated much more than low-frequency voltages: the coupling coefficient for FS action potentials, when measurable, was 0.010 ± 0.006 ($n = 14$). We estimate that the coupling conductance between FS cells was 1.6 ± 1.3 nS (range of 0.4–5.5 nS, $n = 24$; Fig. 3b). Coupling strength was roughly equal in both directions, and showed no dependence on junctional voltage over ± 40 mV. The coupling conductance accounted for up to one-third of the input conductance for the most strongly coupled FS cell pairs. The filtering characteristics of FS–FS cell coupling were tested with subthreshold current steps or sine waves of varying frequencies; coupling was strongest for steady voltages, and rolled off (decreased steeply) at high frequencies with a corner frequency of about 10 Hz. Among

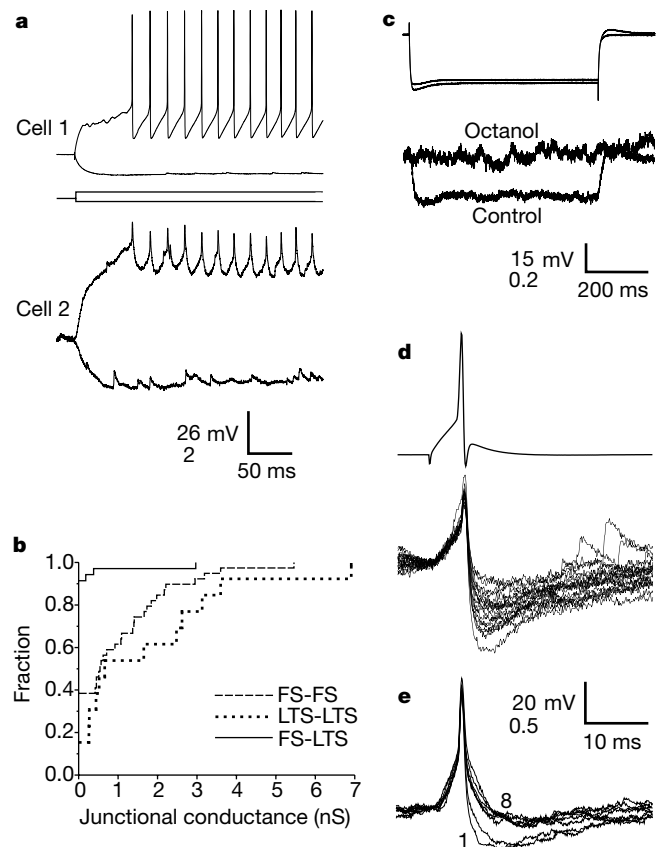


Figure 3 Electrical synapses between interneurons. **a**, Simultaneous recordings from two electrically coupled FS cells. Sequential current steps were applied to Cell 1 only (+400 pA, –200 pA). **b**, Cumulative frequency histogram of junctional conductances estimated from FS–FS ($n = 39$), LTS–LTS ($n = 13$) and FS–LTS ($n = 32$) neuron pairs. **c**, Octanol blocks electrical coupling. Average traces from two coupled FS cells under control conditions and in the presence of octanol (2 mM). Test current steps (–500 pA) were applied to the upper cell. **d**, Dual electrical–chemical synaptic connection between two FS cells. Spikes evoked in Cell 1 with short current pulses (single trace) triggered brief, sharp spikelets in Cell 2 followed by more prolonged and variable IPSPs (multiple trials overlaid). Postsynaptic responses were aligned on the presynaptic spike. **e**, Differential frequency sensitivity of electrical and chemical synaptic components. In an FS–FS pair (different from the pair in **d**), a 10-Hz train of 8 presynaptic spikes (not shown) evoked fast, reliable postsynaptic spikelets by electrical coupling, followed by IPSPs that were strongly depressed after the first 2 stimuli (first and last traces are numbered; each trace is the average of 15 trials).

LTS cell pairs, 85% were electrically coupled (Table 1), with an average coupling coefficient at low frequencies of 0.10 ± 0.09 ($n = 11$), and 0.015 ± 0.01 ($n = 7$) for action potentials. The average junctional conductance between LTS cells was 2.1 ± 2.3 nS (range 0.3–6.9 nS, $n = 11$; Fig. 3b). Strong coupling was observed at every age tested, including one pair of FS cells at postnatal day 26 and one pair of LTS cells at day 29. There was no correlation between junctional conductance and age. Although we demonstrated coupling only between pairs of cells, the prevalence of coupled pairs together with the spatial density of interneurons makes it likely that larger clusters of interneurons are electrically coupled. Electrical synapses were common only between interneurons of the same type (Table 1). Only 3 out of 32 mixed-type (FS–LTS) cell pairs were coupled, and 2 of those had very low junctional conductances (Fig. 3b). Furthermore, we never observed RS cells to be coupled to other RS cells or to interneurons.

Electrical coupling between neurons is usually mediated by gap junctions, which are clusters of intercellular channels formed from connexin proteins¹⁹. Alternative explanations, such as ephaptic or electrical field interactions²⁰, are very unlikely in this case because signal polarity was preserved and no transient capacitive coupling was observed. When electrically coupled interneurons were exposed to the gap junction blocker octanol²¹ (1–2 mM), coupling was abolished in all cases (Fig. 3c; four FS–FS pairs, one LTS–LTS pair), and coupling recovered after washout of octanol in three of

these. Octanol also nonspecifically suppressed action potentials, increased voltage noise and changed cell input resistance, but these effects did not account for the abolition of coupling.

Gap junctions often mediate intercellular chemical signalling¹⁹, and dye coupling (the intercellular passage of a dye of low molecular weight) is often used as an indicator of gap junctions. We filled 24 interneurons with the dye neurobiotin. Although the dye-injected cells were all darkly stained, only one (an FS cell) had a clearly stained interneuron dye coupled to it. Dissociation between dye coupling and electrical coupling may have occurred because the diffusion rate of the dye through small, distally located gap junctions was too slow to allow the coupled cells to be seen. Alternatively, interneurons may be coupled through subtypes of connexins that form channels impermeable to the commonly used dyes²². A variety of connexins exist in the neocortex²³, but the connexin expression patterns of specific neuron types are not known. It is unlikely that the electrical coupling of interneurons is related to the developmental decline in dye coupling, primarily in excitatory cells, that has been observed^{13,14}.

Many FS cell pairs displayed both an electrical synapse and a chemical inhibitory synapse in at least one direction (Table 1). In such pairs, a spike in one cell evoked a biphasic voltage change in the coupled cell: a brief, short-latency, reliable depolarization communicated by the gap junction, and a slower, longer-latency, fluctuating IPSP triggered by the GABA-mediated synapse (Fig. 3d). Because the inhibitory synapses showed short-term depression and electrical connections did not, spike trains (0.5–40 Hz) in the first cell generated relatively constant depolarizations, but diminishing IPSPs, in the coupled cell (for example, 10 Hz in Fig. 3e). Thus, interneurons can communicate through both electrical and chemical synapses that display distinct polarities, timing and short-term dynamics.

Electrical coupling may serve to synchronize the electrical activity of neurons, as it does in the locus coeruleus²⁴ and inferior olive²⁵. Interneurons of the cerebral cortex display various modes of synchronous and rhythmic activity, and both chemical and electrical synapses have been proposed as mediators of their synchrony^{26,27}. We tested the synchronizing ability of electrical coupling by injecting subthreshold sine wave currents (0.5–100 Hz) into one neuron while holding a neighbouring coupled cell at a membrane potential just below spike threshold ($n = 9$, FS–FS pairs). Spiking in the second cell was easily entrained by the depolarizing peaks of the current applied to the first cell at frequencies up to 40 Hz (Fig. 4a). Alternatively, two coupled cells were induced to fire action potentials repetitively (50–100 Hz) with constant current injections. Chemical synapses were either blocked with drugs, or did not exist between the recorded neurons. In all cases (four FS–FS pairs, one LTS–LTS pair) sharply synchronous firing occurred (Fig. 4b).

Electrical coupling may contribute to the markedly sharp (± 1 ms) spiking synchrony observed between pairs of FS cells *in vivo*¹⁰. Close synchrony among electrically coupled LTS cells is also predicted. The paucity of electrical connections between networks, however, should ensure that the fine patterns of FS and LTS spiking are independent. Some coordination between the two networks may occur through their reciprocal inhibitory connections. Several studies suggest that interneurons generate a variety of synchronous inhibitory rhythms in the neocortex^{15,27,28}. Some of this variety may arise from the independent actions of electrically coupled LTS and FS networks, which are regulated by their differential sensitivity to neuromodulators^{29,30}.

Methods

Slice preparation and recording

Thalamocortical slices 250–450 μ m thick were obtained from Sprague–Dawley rats aged P14–P21 (unless otherwise noted). Slices were incubated at 32 °C for 1 hr, and then held at 32 °C during recordings. The bathing solution contained 126 mM NaCl, 3 mM KCl,

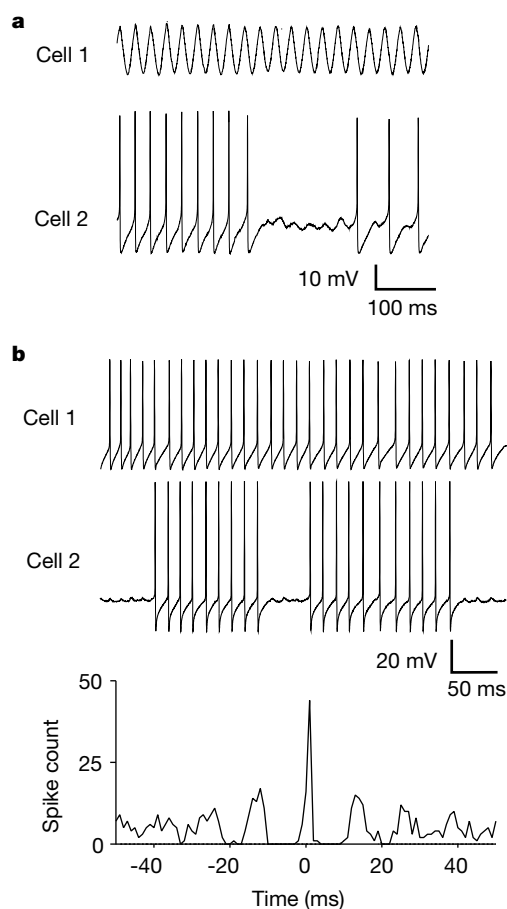


Figure 4 Electrical synapses coupling the spiking activity of two interneurons. **a**, During paired recordings of two FS cells, sinusoidal stimulus current (40 Hz) was applied to Cell 1 while Cell 2 was held just below spike threshold with constant current. Cell 2 fired phasically during many of the positive current peaks in Cell 1. **b**, In a different pair of FS cells, spikes were elicited by simultaneously applying steady depolarizing current to both cells. The traces represent a sample of a longer trial from which a cross-correlation of action potentials was calculated. Cell 1 was the reference for the correlation. The cross-correlogram below shows strong, sharp (± 2 ms) synchrony.

1.25 mM NaH₂PO₄, 2 mM MgSO₄, 26 mM NaHCO₃, 10 mM dextrose and 2 mM CaCl₂, saturated with 95% O₂/5% CO₂. Micropipettes were filled with 135 mM K-gluconate, 4 mM KCl, 2 mM NaCl, 10 mM HEPES, 0.2–4 mM EGTA, 4 mM ATP-Mg, 0.3 mM GTP-Tris, 0.5–10 mM phosphocreatine-Tris (pH 7.25, 295 mOsm). All recordings were made in current clamp mode, with infrared-differential interference contrast (IR-DIC) visualization using a Zeiss Axioskop and a CCD camera (Hamamatsu).

Extracellular stimuli were biphasic current pulses lasting 200 μ s, typically at about 50 μ A (10–120 μ A), applied through microwires. Thalamic responses were measured at membrane potentials of –60 to –70 mV, with 50 μ M DL-2-amino-5-phosphopentanoic acid (AP5) in the bath to block N-methyl-D-aspartate (NMDA) receptors. In a few experiments (see Table 1) 20 μ M of the AMPA receptor antagonist 6,7-dinitroquinoxaline-2,3-dione (DNQX) was also added.

Thalamocortical responses were collected with a conservative 'minimal response' protocol: threshold stimulus intensities evoked EPSPs (up to 100 trials) presumed to be generated by single thalamocortical axons. Most EPSP failures probably arose from failure of presynaptic axon spiking; incrementing stimulus intensity abolished failures without increasing EPSP size and shape. Thalamocortical responses were discarded from cells in which IPSPs were detected.

Junctional conductance

Estimating the junctional conductance¹⁸ assumed a model of two isopotential neurons and a single junction. Junctional conductance and the two cell input conductances were solved from simultaneous equations, using the injected currents and voltage responses of each cell. Coupling conductances were estimated for each direction of current flow in a cell pair; differences were usually small, and reported values are the average of the two. The model does not account for complexities arising from junctions on dendritic or axonal cables, and it does not incorporate nonlinear membrane properties.

Anatomical labelling

Neurobiotin or biocytin (4 mg ml^{–1}) was added to the normal filling solution in some recordings. Slices were fixed in 4% paraformaldehyde (sometimes with 0.2% picric acid) in 0.1 M phosphate buffer, transferred to 30% sucrose, resectioned to 80 μ m, and reacted with avidin-biotin-peroxidase (Vector). In dye-coupling experiments with neurobiotin, cells were recorded for 1–1.75 hr and, in about half of the cells, current steps were applied in an effort to maximize dye delivery.

Some cells underwent a fluorescent double-labelling immunocytochemical protocol for biocytin and either parvalbumin or somatostatin, using two different fluorophores. Slices were first stained using monoclonal antibodies for either parvalbumin (Sigma, S-3171, diluted 1:400, visualized by CY₃-conjugated goat anti-mouse IgG (Jackson Immuno Research), or somatostatin (S-309, gift of R. Benoit, Montreal, Canada, visualized by CY₃-conjugated donkey anti-rabbit IgG (Chemicon). Biocytin was visualized using 7-amino-4-methylcoumarin-3-acetic acid (AMCA)-conjugated streptavidin (Jackson Immuno Research). Immunonegative cells could result from genuine scarcity of the marker in a cell, or could reflect damage to a recorded cell.

Received 18 May; accepted 26 August 1999.

- Somogyi, P., Tamas, G., Lujan, R. & Buhl, E. H. Salient features of synaptic organisation in the cerebral cortex. *Brain Res. Rev.* **26**, 113–135 (1998).
- Jones, E. G. GABAergic neurons and their role in cortical plasticity in primates. *Cerebral Cortex* **3**, 361–372 (1993).
- Agmon, A. & Connors, B. W. Thalamocortical responses of mouse somatosensory (barrel) cortex in vitro. *Neuroscience* **41**, 365–379 (1991).
- Simons, D. J. & Woolsey, T. A. Morphology of Golgi-Cox-impregnated barrel neurons in rat SmI cortex. *J. Comp. Neurol.* **230**, 119–132 (1984).
- Thomson, A. M. & Deuchars, J. Synaptic interactions in neocortical local circuits: dual intracellular recordings in vitro. *Cereb. Cortex* **7**, 510–522 (1997).
- Kawaguchi, Y. & Kubota, Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb. Cortex* **7**, 476–486 (1997).
- Reyes, A. & Sakmann, B. Target-cell-specific facilitation and depression in neocortical circuits. *Nature Neurosci.* **1**, 279–285 (1998).
- Connors, B. W. & Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13**, 99–104 (1990).
- Agmon, A. & Connors, B. W. Correlation between intrinsic firing patterns and thalamocortical synaptic responses of neurons in mouse barrel cortex. *J. Neurosci.* **12**, 319–329 (1992).
- Swadlow, H. A., Beloozerova, I. N. & Sirota, M. G. Sharp, local synchrony among putative feed-forward inhibitory interneurons of rabbit somatosensory cortex. *J. Neurophysiol.* **79**, 567–582 (1998).
- Staiger, J. F., Zilles, K. & Freund, T. F. Distribution of GABAergic elements postsynaptic to ventroposteromedial thalamic projections in layer IV of rat barrel cortex. *Eur. J. Neurosci.* **8**, 2273–2285 (1996).
- Gil, Z., Connors, B. W. & Amitai, Y. Efficacy of thalamocortical and intracortical synaptic connections: quanta, innervation, and reliability. *Neuron* **23**, 385–397 (1999).
- Connors, B. W., Bernardo, L. S. & Prince, D. A. Coupling between neurons of the developing rat neocortex. *J. Neurosci.* **3**, 773–782 (1983).
- Peinado, A., Yuste, R. & Katz, L. C. Extensive dye coupling between rat neocortical neurons during the period of circuit formation. *Neuron* **10**, 103–114 (1993).
- Benardo, L. S. Recruitment of GABAergic inhibition and synchronization of inhibitory interneurons in rat neocortex. *J. Neurophysiol.* **77**, 3134–3144 (1997).
- Sloper, J. J. & Powell, T. P. Gap junctions between dendrites and somata of neurons in the prime sensori-motor cortex. *Proc. R. Soc. Lond. B* **203**, 39–47 (1978).
- Bozhilova-Pastirova, A. & Ovtcharoff, W. Structure of the synaptic junctions in the rat sensorimotor cortex: freeze-etching study of neuronal gap junctions. *Neurosci. Lett.* **201**, 265–267 (1995).
- Bennett, M. V. L. in *Handbook of Physiology, Section 1: The Nervous System, Part I* (eds Brookhart, J. M. & Mountcastle, V. B.) 357–416 (American Physiological Society, Bethesda, 1977).

- Dermietzel, R. & Spray, D. C. Gap junctions in the brain: where, what type, how many and why? *Trends Neurosci.* **16**, 186–192 (1993).
- Jefferys, J. G. R. Nonsynaptic modulation of neuronal activity in the brain: electric currents and extracellular ions. *Physiol. Rev.* **75**, 689–723 (1995).
- Johnston, M. E., Simon, S. A. & Ramon, F. Interaction of anesthetics with electrical synapses. *Nature* **286**, 498–500 (1980).
- Veenstra, R. D. Size and selectivity of gap junction channels formed from different connexins. *J. Bioenerg. Biomembr.* **28**, 327–337 (1996).
- Nadarajah, B., Jones, A. M., Evans, W. H. & Parnevelas, J. G. Differential expression of connexins during neocortical development and neuronal circuit formation. *J. Neurosci.* **17**, 3096–3111 (1997).
- Christie, M. J., Williams, J. T. & North, R. A. Electrical coupling synchronizes subthreshold activity in locus coeruleus neurons *in vitro* from neonatal rats. *J. Neurosci.* **9**, 3584–3589 (1989).
- Mann-Metzer, P. & Yarom, Y. Electrotonic coupling interacts with intrinsic properties to generate synchronized activity in cerebellar networks of inhibitory interneurons. *J. Neurosci.* **19**, 3298–3306 (1999).
- Michelson, H. B. & Wong, R. K. Synchronization of inhibitory neurones in the guinea-pig hippocampus *in vitro*. *J. Physiol. (Lond.)* **477**, 35–45 (1994).
- Jefferys, J. G., Traub, R. D. & Whittington, M. A. Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208 (1996).
- Buhl, E. H., Tamas, G. & Fisahn, A. Cholinergic activation and tonic excitation induce persistent gamma oscillations in mouse somatosensory cortex *in vitro*. *J. Physiol. (Lond.)* **513**, 117–126 (1998).
- Kawaguchi, Y. & Shindou, T. J. Noradrenergic excitation and inhibition of GABAergic cell types in rat frontal cortex. *J. Neurosci.* **18**, 6963–6976 (1998).
- Xiang, Z., Huguenard, J. R. & Prince, D. A. Cholinergic switching within neocortical inhibitory networks. *Science* **281**, 985–988 (1998).

Acknowledgements

We thank S. Patrick for technical assistance, Y. Amitai and M. Bear for comments on the manuscript, and R. Benoit for gifts of anti-somatostatin antibodies. This work was supported by a fellowship to J.R.G. from NIH and a grant to B.W.C. from NIH.

Correspondence and request for materials should be addressed to B.W.C. (e-mail: Barry_Connors@Brown.edu).

Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice

Jeffery S. Cox*, Bing Chen*, Michael McNeil† & William R. Jacobs Jr*

* Howard Hughes Medical Institute, Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA
† Department of Microbiology, Colorado State University, Ft. Collins, Colorado 80523, USA

Tuberculosis is the leading cause of death in the world resulting from a single bacterial infection¹. Despite its enormous burden on world health, little is known about the molecular mechanisms of pathogenesis of *Mycobacterium tuberculosis*. Bacterial multiplication and concomitant tissue damage within an infected host, including experimentally infected mice, occurs primarily in the lungs—the favoured niche of *M. tuberculosis*². Although it has been proposed that the distinctive cell wall of *M. tuberculosis* is important for virulence, rigorous genetic proof has been lacking. Here, using signature-tagged mutagenesis, we isolated three attenuated *M. tuberculosis* mutants that cannot synthesize or transport a complex, cell wall-associated lipid called phthiocerol dimycocerosate (PDIM) which is found only in pathogenic mycobacteria^{3,4}. Two mutants have transposon insertions affecting genes implicated in PDIM synthesis; the third has a disruption in a gene encoding a large transmembrane protein required for proper subcellular localization of PDIM. Synthesis and transport of this complex lipid is only required for growth in the lung; all three mutants are unaffected for growth in the liver and spleen. This clearly shows that a lipid is required for *M. tuberculosis* virulence.

To identify *M. tuberculosis* genes that are required for growth *in vivo*, we adopted the signature-tag mutagenesis (STM)⁵ scheme to screen multiple 'inoculum' pools of 48 random transposon mutants

for variants that were unable to replicate within mouse lungs. Because each transposon carried a unique sequence tag, the individual clones within each pool could be distinguished from one another by detection of these tags using hybridization. We looked for mutants containing tags that were abundant in the inoculum pool but were scarce or absent after growth *in vivo*. As an initial screen, we created 12 sets of mutant pools and used each set to infect two immunocompetent C57BL/6 mice. After three weeks, mycobacteria were harvested from the lungs and DNA tags were amplified and used as probes on separate tag arrays. Figure 1a shows the result of a representative pool in which all of the 48 mutants were present in the inoculum pool (Fig. 1a, 'inoculum') but two mutants were significantly under-represented from the lungs of each of the infected mice (Fig. 1a, 'in vivo selected'). Only mutants containing tags that were under-represented or missing in both of the infected mice were selected for further study. In total, 14 *giv* (growth *in vivo*) mutants were isolated out of 576 transposon-containing isolates screened, a rate (2.4%) similar to that found in STM screens of other bacterial pathogens (0.5–4.0%)^{5–7}.

The insertion site of the transposon within the genome of each *giv* mutant was determined and compared with the genomic sequence of *M. tuberculosis*. Most of the transposon insertions disrupted open reading frames encoding proteins of unknown function (data not shown). We therefore focused our initial studies on three *giv* mutants with insertions within a 44-kilobase (kb) region containing genes whose products are involved in the synthesis of PDIM.

The first mutant contains a transposon insertion in the promoter region of a large operon that includes the *ppsA-E* genes (Fig. 1b). These genes are highly homologous to those encoding the multi-subunit polyketide synthase required for phthiocerol biosynthesis in *M. bovis* BCG⁸. Immediately downstream of the *pps* operon is the *M. tuberculosis* homologue of the *mas* gene which in BCG encodes a fatty acid synthase-like enzyme responsible for the synthesis of mycocerosic acids^{9,10}. The products of these two synthases are attached covalently by an unknown mechanism to form PDIM (Fig. 2b).

The second *giv* mutant contains an insertion in *fadD28*, one of 36 homologues of the *Escherichia coli* *fadD* gene identified in the

M. tuberculosis genome¹¹. *FadD28*, however, is probably involved in acyl transfer of mycocerosic acid instead of fatty acid catabolism, as a mutation of the *fadD28* homologue in BCG specifically inhibits mycocerosic acid synthesis and thus PDIM synthesis^{12,13}.

Finally, the third *giv* mutant has a transposon insertion in the *mmpL7* gene, which encodes a putative transmembrane protein with high homology to a number of proteins encoded by the *M. tuberculosis* genome as well as ActIII-ORF3, a membrane protein required for secretion of the antibiotic actinorhodin in *Streptomyces coelicolor*¹⁴. The *mmpL7* mutant, like the *pps* and *fadD28* mutants, is specifically defective for *in vivo* growth, as its doubling time in liquid culture is identical to that of the parent Erdman strain (not shown).

Our first indication that all three attenuated mutants were defective for cell-wall biosynthesis was that each displayed a strikingly altered colony morphology as compared with wild-type *M. tuberculosis* when grown on solid medium (Fig. 1c). Whereas wild-type cells formed characteristically flat 'corded' colonies, each of the mutant cells produced colonies of overlapping tubular structures which rose above the plane of the agar plate. Colony morphology in *M. tuberculosis* has long been associated with virulence^{15,16}. Although surface-exposed lipids have been implicated as key contributors to this phenotype, this is the first demonstration of a link between a specific lipid biosynthetic pathway, colony morphology and virulence.

Given that all the genes in this region of the *M. tuberculosis* genome are nearly identical to those in BCG, we asked whether these transposon mutants were deficient for PDIM synthesis. PDIM can be specifically labelled *in vivo* with ¹⁴C-propionate, as it is a precursor to methylmalonyl-coenzyme A, the substrate used by these synthases to introduce methyl branches onto fatty acids¹⁷. Therefore, to determine whether the three transposon mutants can produce PDIM, cells were labelled with ¹⁴C-propionate, and whole-cell lipids were extracted and analysed by thin-layer chromatography (TLC). As shown in Fig. 2a, extracts prepared from wild-type *M. tuberculosis* cells contained substantial amounts of two radiolabelled lipids (lane 1). The lipid of higher mobility (lipid 1, *R_f* = 0.69) is probably PDIM, as it co-migrated with a well characterized PDIM standard purified from *M. leprae*; the species of lower mobility (lipid 2, *R_f* = 0.55) may be related to PDIM, owing to incorporation of ¹⁴C-propionate. Importantly, extracts prepared from the *pps*-promoter mutant and the *fadD28* mutant contained little or no detectable amount of either lipid (Fig. 2a, lanes 2 and 3). In contrast, *mmpL7* mutant cells accumulated both of the labelled lipids, although the relative amounts of the two species were altered as compared with the wild-type strain, with more lipid 1 present (Fig. 2a, lane 4).

To verify the identity of the two lipid species missing in the *pps*-promoter and the *fadD28* mutants, lipids corresponding to the two bands on the chromatograph were purified from wild-type *M. tuberculosis* and subjected to mass spectral analysis. During fast atom bombardment/mass spectrometry (FAB/MS) one of the two mycocerosic acid residues is readily lost; the structure corresponding to this ion is shown in Fig. 2b. In the spectrum obtained from the standard PDIM from *M. leprae* (Fig. 2c, top), the *m/z* 888 peak corresponds exactly to the form of PDIM in which the base chain lengths of the phthiocerol and mycocerosic acid subunits are 28 and 27 carbon atoms long, respectively, and is consistent with the reported structure of this molecule¹⁸. The higher *m/z* peaks in this spectrum occur at intervals of 14 atomic mass unit (a.m.u.), which corresponds to heterogeneity in the chain length (CH₂ units) of the mycocerosic acid and the phthiocerol moieties, a phenomenon consistently observed for this class of molecules¹⁸. An analogous spectrum was obtained from lipid 1 except that major ions were greater than those from *M. leprae* PDIM, again by multiples of 14 a.m.u., consistent with the fact that the base chain of phthiocerol is longer in *M. tuberculosis*¹⁹ (Fig. 2c, middle). Thus, lipid 1 is indeed PDIM with the *m/z* 930 peak representing the molecule drawn in

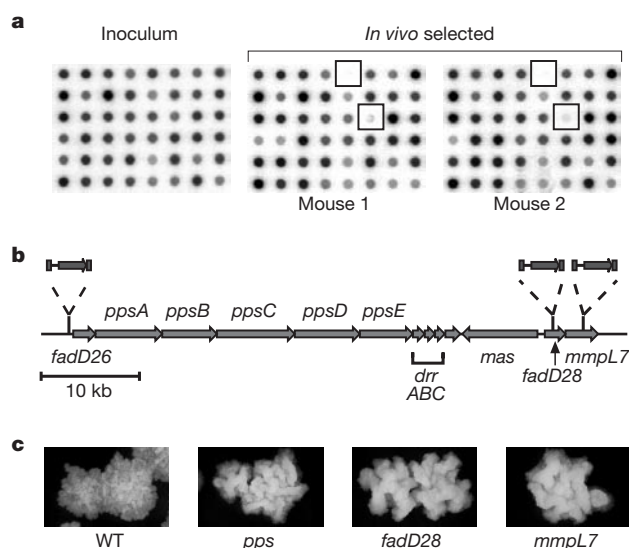


Figure 1 Three related, avirulent mutants of *M. tuberculosis* isolated by STM mutagenesis. **a**, Signature tags from mycobacteria harvested from the inoculum and the lungs of mice after 3 weeks of growth were amplified, radiolabelled and hybridized to tag-array filters. Results from one pool are shown and tags from mutants under-represented *in vivo* are boxed. **b**, Schematic representation of the transposon insertion sites from three *giv* mutants within the *M. tuberculosis* genome. **c**, Wild-type (WT) *M. tuberculosis* and the *pps*-promoter, *fadD28* and *mmpL7* mutants were plated for single colonies on solid medium and incubated for 4 weeks at 37 °C.

Fig. 2b, where R is OCH₃ (this mass also fits other possible combinations of sizes of the phthiocerol and mycocerosates). As expected, the lipid of lower mobility (lipid 2) is also a PDIM-like compound but yielded fragments 16 a.m.u. lower than *M. tuberculosis*

losis PDIM (Fig. 2c, bottom). This pattern is expected if lipid 2 contains phthiodiolone instead of phthiocerol, which differs from PDIM only by the replacement of the phthiocerol methoxy group with a keto group (Fig. 2b, R is O). This idea was confirmed by ¹H-nuclear magnetic resonance analysis in which a methoxy peak at 3.95 p.p.m. was present in PDIM but absent in lipid 2 (not shown). Thus, we conclude that lipid 2 is phthiodiolone dimycocerosate, the presumed biosynthetic precursor of PDIM¹⁹.

Initially, the observation that *mmpL7* mutant cells could synthesize PDIM but displayed all the phenotypes (inability to grow *in vivo*, altered colony morphology) of the other two STM mutants, which produced no PDIM at all, was puzzling. However, the apparent accumulation of mature PDIM in the mutant and the homology between *mmpL7* and *actII-ORF3* genes suggested that MmpL7 may be involved in PDIM transport. To verify this proposal, we performed pulse-chase labelling experiments with ¹⁴C-propionate to follow the fate of newly synthesized PDIM in wild-type and *mmpL7* mutant cells. A considerable portion of the labelled PDIM is found in the liquid culture medium from wild-type *M. tuberculosis* cells 16 h after synthesis (Fig. 2d). In contrast, *mmpL7* mutant cells were unable to liberate PDIM into the culture medium, and all of the labelled lipid remained associated with the cell pellet (Fig. 2d, lower panels). In three separate experiments, wild-type cultures released an average of 45% of the labelled PDIM into the medium during the 16-h chase period, whereas PDIM was never

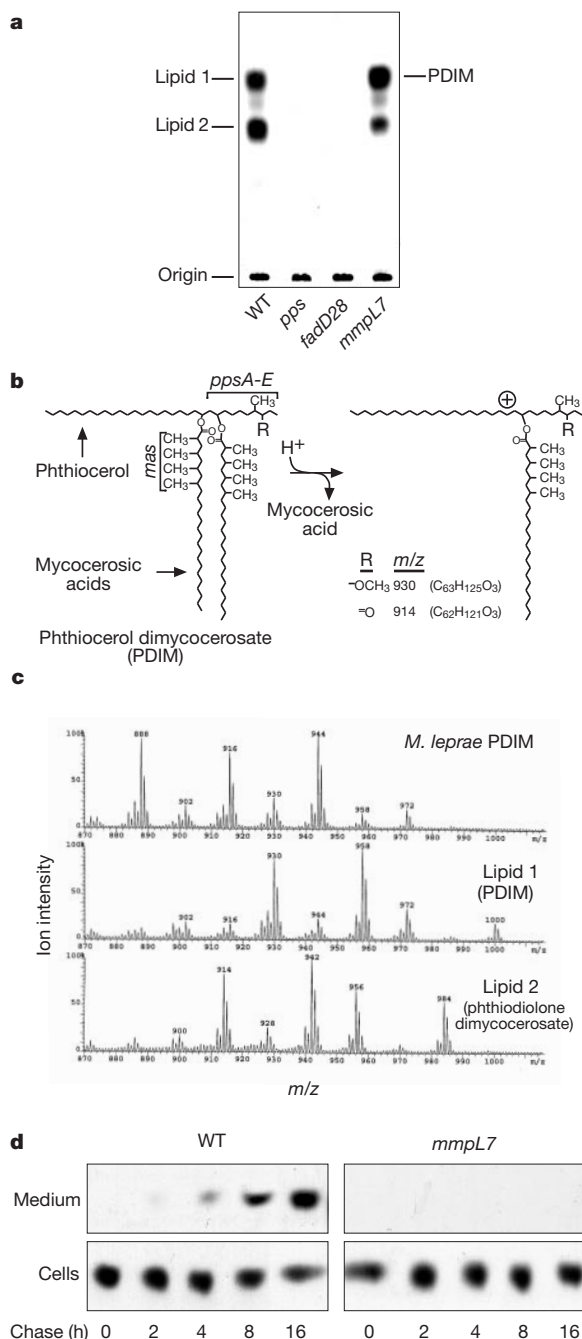


Figure 2 The three *giv* mutants cannot synthesize or transport PDIM. **a**, Cells were labelled with [¹⁴C]propionic acid for 16 h and crude lipid extracts were separated by TLC. The position of authentic, purified PDIM from *Mycobacterium leprae* is denoted. **b**, Left, model of PDIM structure showing sites of propionic acid incorporation by the action of the *ppsA-E* and *mas* gene products⁴. Right, formation of the major diagnostic PDIM ions during FAB/MS. Exact mass numbers, rounded to the nearest whole number, are given for ions containing phthiocerol (methoxy) and phthiodiolone (keto). **c**, Spectra of purified *M. leprae* PDIM and *M. tuberculosis* lipids 1 and 2 from **a**. **d**, Pulse-chase analysis. Wild-type (WT) and *mmpL7* mutant cells were labelled with [¹⁴C]propionic acid for 2 h, washed and re-incubated in fresh medium. Lipid extracts from cell pellets and filtered medium were isolated during the chase period as described in Methods. TLC plates were developed with chloroform:methanol (19:1) in which PDIM and phthiodiolone dimycocerosate co-migrated as a single species.

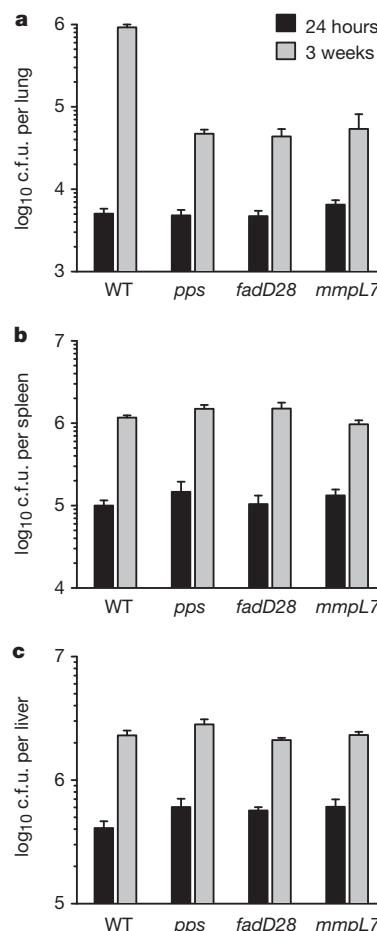


Figure 3 PDIM synthesis and export is required for *M. tuberculosis* replication in mouse lungs but not in the liver or spleen. **a**, C57BL/6 mice were infected with 1 × 10⁶ c.f.u. of each strain, and *M. tuberculosis* cells were harvested from lungs at 24 h (black bars) and 3 weeks (grey bars) after infection and counted by plating. Error bars represent the standard error from two separate experiments using at least three infected mice for each time point per experiment. **b**, **c**, *M. tuberculosis* cells were harvested and counted from spleens (**b**) and livers (**c**) from the mice in **a**.

detected in the culture medium from the *mmpL7* mutant. Although we cannot currently distinguish whether PDIM is actively secreted by wild-type *M. tuberculosis* or simply sloughed off into the medium owing to its position on the cell surface, it is clear that MmpL7 is at least required for PDIM export to the cell wall.

The isolation of these three independent mutants from the STM screen strongly suggested that PDIM production and localization by *M. tuberculosis* is important for growth in the lungs of the host. However, as these mutants were isolated from pools of infected bacteria, we sought to verify that they were also defective for growth when injected as a clonal population. Therefore, we infected C57BL/6 mice with 1×10^6 colony-forming units (c.f.u.) of either the wild-type Erdman strain or each of the three transposon mutants. Wild-type *M. tuberculosis* cells could replicate extensively in the lungs during the first three weeks after infection, leading to nearly a 200-fold increase in the number of viable mycobacteria recovered from the organ (Fig. 3a, WT). However, in mice infected with cells from either the *pps*-promoter, *fadD28* or *mmpL7* mutant strains, mycobacterial replication was severely restricted in the lungs (Fig. 3a). In agreement with their ability to control replication of the mutants, the mice infected with the attenuated strains exhibited less severe pneumonia than mice infected with wild-type *M. tuberculosis* (see Fig. 1 in Supplementary Information). Note, however, that although all three mutants were compromised for growth in the lung as compared with the wild type, they were still able to grow slightly in the organ (~8-fold increase).

Although replication of wild-type *M. tuberculosis* is greatest in the lung, infection by intravenous injection also leads to the seeding of mycobacterial cells into other organs where they can multiply. To determine the capacity of the three *giv* mutants to grow in tissues other than the lung, we monitored the number of mycobacteria in the spleens and livers of the same infected mice used for the experiments described above. All three mutants had growth rates comparable with the wild type in both organs (Fig. 3b, c). These results show that PDIM is specifically required for growth of *M. tuberculosis* in the lungs of infected mice.

PDIM synthesis is complex and requires the action of the polyketide synthase-like Pps enzymes, Mas, at least one acyl-CoA synthase-like protein and MmpL7 (Fig. 4). As the *mmpL7* mutation affects PDIM export, we believe that MmpL7 is likely to be involved in the direct transfer of PDIM across the cytoplasmic membrane. It will therefore be interesting to determine whether PDIM acts solely on the cell surface of *M. tuberculosis*, perhaps as a mediator of cell-wall permeability, or whether PDIM secretion into the host is important for virulence. Finally, the presence of 11 other *mmpL*

genes in the genome of *M. tuberculosis* suggests that this pathogen may export a diverse set of lipids involved in pathogenesis. □

Methods

Media and Strains

M. tuberculosis strains were grown in 7H9 liquid medium supplemented with 10% OADC, 0.5% glycerol and 0.1% Tween-80, or in 7H10 solid medium with the same supplements except for Tween-80. STM medium is either 7H9 or 7H10 supplemented with 0.5% casamino acids, $20 \mu\text{g ml}^{-1}$ tryptophan, 10% OADC, 0.5% glycerol, $1 \mu\text{g ml}^{-1}$ cyclohexanamide and $50 \mu\text{g ml}^{-1}$ hygromycin B. The *pps* promoter, *fadD28* and *mmpL7* mutant strains are named mc²3105, mc²3106 and mc²3107 respectively. The wild-type Erdman strain carrying the integrating vector pMV306 is named mc²3104.

STM mutagenesis

The random tag library was constructed and amplified as described⁵ except that an *SphI* site was introduced at the ends of the tags. Tags were ligated into the unique *SphI* site of the Tn5370 transposon in pJSC84 and the resulting plasmids were introduced into the genome of the temperature-sensitive phage phAE87 (ref. 20). Forty-eight phages containing tags that hybridized well to their cognate tags and not to the other 47 tags were selected. Tag arrays were made by applying either amplified tags or phage genomic DNA onto hybond-N⁺ nylon filters (Amersham Pharmacia Biotech) using a Bio-Dot filter apparatus (Bio-Rad). Transposon mutagenesis²⁰ of the Erdman strain was performed using all 48 phages separately. Colonies were picked into 48-well plates and portions of the cultures were combined to create the inoculum pool, which was frozen and titred. 6–8 week old female C57BL/6 mice (Jackson Labs) were infected by tail vein injection with $\sim 3 \times 10^6$ c.f.u., and the lungs were removed and homogenized three weeks after infection. Total DNA from the homogenate was prepared by boiling and bead-beating in the presence of phenol. The insertion site of the transposon in the selected mutants was determined by sequencing the products of inverse PCR²¹. Briefly, 0.5 μg of genomic DNA was digested with *RsaI* and ligated to form circles, and then genomic sequences flanking both sides of the transposon were amplified. Primers used to amplify from the left side of the transposon were o84L-F (5'-GTCATCCGGCTCATCACCAG-3') and o84L-R (5'-AACTGGCG-CAGTTCCTCTCTGG-3'), and the right side was amplified using o84R-F (5'-ATACACGCGCACCGGTTCTAGC-3') and o84R-R (5'-CACGCGCAACGCTGGTG-3'). Finally, DNA from each mutant was subjected to Southern analysis using the hygromycin resistance marker gene as a probe to certify that only one transposon was present in the genome.

Biochemical analysis of PDIM

Labelling, extraction and analysis of lipids were done essentially as described⁸. Actively growing cultures were diluted to OD₆₀₀ = 0.8 in a final volume of 50 ml. $20 \mu\text{Ci}$ of Na[1-¹⁴C]propionate (American Radiolabeled Chemicals) was added and incubation continued for 16 h at 37 °C. Cell pellets were washed twice with water and then extracted with 5 ml of chloroform:methanol (2:1) at room temperature overnight. Lipids were prepared by the Folch method²², dried under nitrogen and dissolved in 1 ml of ethyl ether. 4 μl was separated on 10 cm $\times$ 10 cm HPTLC plates (Alltech) using either hexanes:ether (9:1), which separated PDIM from phthiodiolone dimycocerosate, or chloroform:methanol (19:1), in which both lipids co-migrated, as the solvent. 5 μg purified PDIM (gift from P. Brennan) was applied as a marker and visualized by charring. Lipids were isolated from preparative TLC plates and eluted with ethyl ether. Mass spectrometry was performed on a Fisons VG autospec FAB with a caesium ion source using m-nitrobenzyl alcohol as the matrix. For pulse-chase experiments, 80 ml of culture was incubated with 32 μCi of Na[1-¹⁴C]propionate for 2 h. Cells were centrifuged, washed with fresh medium three times and resuspended in 80 ml medium. 15-ml aliquots were removed at different time points and cells were pelleted, washed and extracted. Medium was filtered through 0.2 μm filters, lyophilized and then extracted as done for cell pellets.

Infection of individual STM mutants

Actively growing cultures of mc²3104–3107 were washed, resuspended in $1 \times$ PBS and 0.1% Tween-80 and sonicated. C57BL/6 mice were infected by tail vein injection with 1×10^6 c.f.u. of each strain using an OD₆₀₀ = 3×10^8 c.f.u. ml⁻¹. Organs from infected mice were homogenized and plated as described²³.

Received 2 September; accepted 30 September 1999.

1. World Health Organization, World Health Report 1999 <http://www.who.int/whr/1999/en/report.htm> (1999).
2. Garay, S. M. in *Tuberculosis* (eds Rom, W. N. & Garay, S. M.) 373–412 (Little, Brown and Company, Boston, 1996).
3. Brennan, P. J. & Nikaido, H. The envelope of mycobacteria. *Annu. Rev. Biochem.* **64**, 29–63 (1995).
4. Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**, 263–270 (1997).
5. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
6. Chiang, S. L. & Mekalanos, J. J. Use of signature-tagged transposon mutagenesis to identify *Vibrio cholerae* genes critical for colonization. *Mol. Microbiol.* **27**, 797–805 (1998).
7. Mei, J. M., Nourbakhsh, F., Ford, C. W. & Holden, D. W. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol. Microbiol.* **26**, 399–407 (1997).
8. Azad, A. K., Sirakova, T. D., Fernandes, N. D. & Kolattukudy, P. E. Gene knockout reveals a novel gene

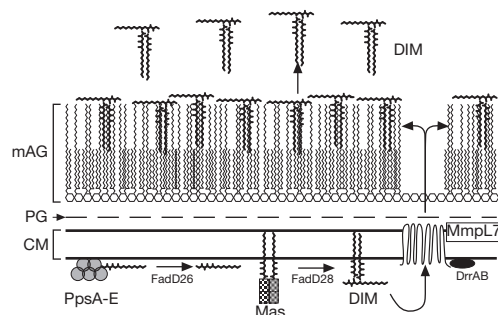


Figure 4 Model for synthesis and export of PDIM. Arrows above FadD26 and FadD28 represent acyl-transfer reactions hypothetically catalysed by these proteins, releasing phthiocerol and mycocerosic acids from their respective synthases (PpsA-E and Mas) and condensing them to form PDIM. PDIM may be transported across the cytoplasmic membrane (CM) by MmpL7—perhaps acting with the ABC transporter-like DrrAB proteins. Presumably PDIM can diffuse through the peptidoglycan (PG) layer and into the lipid-rich mycolyl-arabinogalactan (mAG) layer of the cell wall. Efforts are underway to determine whether release of PDIM from the surface of the bacterium occurs during infection and to identify its role in the host–pathogen interaction.

- cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J. Biol. Chem.* **272**, 16741–16745 (1997).
9. Mathur, M. & Kolattukudy, P. E. Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. bovis Bacillus Calmette-Guerin. *J. Biol. Chem.* **267**, 19388–19395 (1992).
 10. Azad, A. K., Sirakova, T. D., Rogers, L. M. & Kolattukudy, P. E. Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. *Proc. Natl Acad. Sci. USA* **93**, 4787–4792 (1996).
 11. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence [see comments]. *Nature* **393**, 537–544 (1998).
 12. Fitzmaurice, A. M. & Kolattukudy, P. E. An acyl-CoA synthase (acoas) gene adjacent to the mycocerosic acid synthase (mas) locus is necessary for mycocerosyl lipid synthesis in *Mycobacterium tuberculosis* var. bovis. *BCG. J. Biol. Chem.* **273**, 8033–8039 (1998).
 13. Fitzmaurice, A. M. & Kolattukudy, P. E. Open reading frame 3, which is adjacent to the mycocerosic acid synthase gene, is expressed as an acyl coenzyme A synthase in *Mycobacterium bovis* BCG. *J. Bacteriol.* **179**, 2608–2615 (1997).
 14. Bystrykh, L. V. *et al.* Production of actinorhodin-related “blue pigments” by *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **178**, 2238–2244 (1996).
 15. Pierce, C. H. & Dubos, R. J. Differential characteristics *in vitro* and *in vivo* of several substrains of BCG. II. Morphologic characteristics *in vitro* and *in vivo*. *Am. Rev. Tuberc. Pulm. Dis.* **74**, 667–682 (1956).
 16. Middlebrook, G., Dubos, R. J. & Pierce, C. Virulence and morphological characteristics of mammalian tubercle bacilli. *J. Exp. Med.* **86**, 175–183 (1947).
 17. Rainwater, D. L. & Kolattukudy, P. E. Synthesis of mycocerosic acids from methylmalonyl coenzyme A by cell-free extracts of *Mycobacterium tuberculosis* var. bovis BCG. *J. Biol. Chem.* **258**, 2979–2985 (1983).
 18. Hunter, S. W. & Brennan, P. J. Further specific extracellular phenolic glycolipid antigens and a related diacylphthiocerol from *Mycobacterium leprae*. *J. Biol. Chem.* **258**, 7556–7562 (1983).
 19. Besra, G. S. & Chatterjee, D. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 285–306 (American Society for Microbiology, Washington DC, 1994).
 20. Bardarov, S. *et al.* Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **94**, 10961–10966 (1997).
 21. Ochman, H., Gerber, A. S. & Hartl, D. L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623 (1988).
 22. Folch, J., Lees, M. & Stanley, G. H. S. A simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1957).
 23. McAdam, R. A. *et al.* *In vivo* growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* **63**, 1004–1012 (1995).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Brennan for valuable discussions and for purified PDIM; M. Glickman, J. McKinney, R. Morbidoni and P. Draper for helpful discussions; P. Walter, M. Glickman and J. Blanchard for critical review of the manuscript and encouragement; D. Chatterjee, J. Torrelles, D. Dick and M. Scherman for mass spectrum and NMR analysis; R. Russell for valuable discussions and assistance with histopathology; and R. McAdam and S. Quan for the inverse PCR protocol. J.S.C. is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research.

Correspondence and requests for materials should be addressed to W.R.J. Jr (e-mail: jacobs@ucla.edu).

A combined algorithm for genome-wide prediction of protein function

Edward M. Marcotte^{*†}, Matteo Pellegrini^{†‡}, Michael J. Thompson^{*‡}, Todd O. Yeates^{*} & David Eisenberg^{*}

^{*} Molecular Biology Institute, UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, PO Box 951570, Los Angeles, California 90095, USA

[†] These authors contributed equally to this work

The availability of over 20 fully sequenced genomes has driven the development of new methods to find protein function and interactions. Here we group proteins by correlated evolution¹, correlated messenger RNA expression patterns² and patterns of domain fusion³ to determine functional relationships among the 6,217 proteins of the yeast *Saccharomyces cerevisiae*. Using these methods, we discover over 93,000 pairwise links between func-

tionally related yeast proteins. Links between characterized and uncharacterized proteins allow a general function to be assigned to more than half of the 2,557 previously uncharacterized yeast proteins. Examples of functional links are given for a protein family of previously unknown function, a protein whose human homologues are implicated in colon cancer and the yeast prion Sup35.

The historical method of finding the function of a protein involves extensive genetic and biochemical analyses, unless the amino-acid sequence of the protein resembles another whose function is known. With complete genome sequences and total mRNA expression patterns, new strategies become available. We show that the general biochemical functions of proteins can be inferred by associating proteins on the basis of properties other than the similarity between their amino-acid sequences. These properties associate proteins that are functionally related, which we define as proteins that participate in a common structural complex, metabolic pathway, biological process or closely related physiological function. Here we combine three independent methods of prediction^{1–3} with available experimental data^{3–5} to create a large-scale prediction of protein function which does not rely upon direct sequence homology. These methods allow us to establish many thousands of links between proteins of related function in the yeast *S. cerevisiae*⁶.

One might arbitrarily expect each protein coded by a genome to be functionally linked, and therefore to perform closely related functions in the cell, with perhaps 5–50 other proteins, giving 30,000–300,000 biologically meaningful links in yeast. We find 20,749 protein–protein links from correlated phylogenetic profiles¹, 26,013 links from correlated mRNA expression patterns, and 45,502 links from Rosetta Stone sequences of the domain fusion method³.

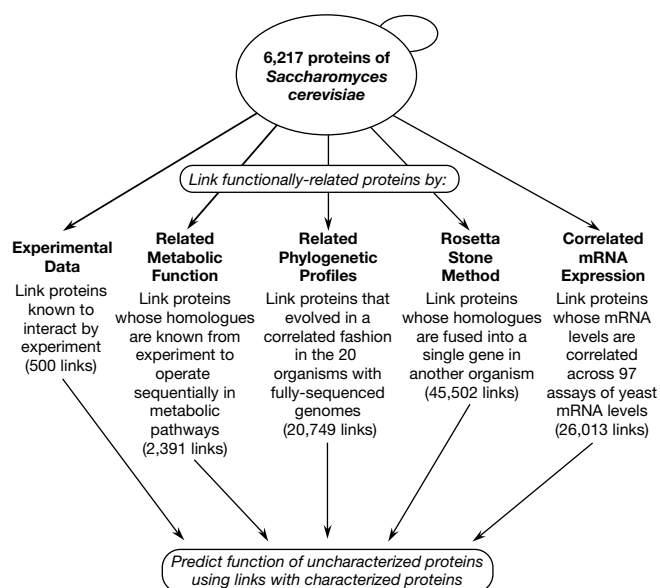


Figure 1 Strategies used to link functionally related yeast proteins, showing the number of links provided by each method. 4,130 very high confidence links were constructed from experimental data, between proteins of related metabolic function, and from links generated by at least two of the prediction methods. 19,521 high confidence links were constructed from the phylogenetic profile method when unconfirmed by other techniques. The remaining links were considered to be of lower confidence. Links associate proteins of related function and can therefore be used to predict the function of many uncharacterized proteins. The phylogenetic profile method¹ identifies functionally related proteins by assuming that proteins that are always inherited together operate together. The fused domain method³ identifies functionally related proteins by assuming that two proteins function together if they appear in some other organism on the same polypeptide chain, called a Rosetta Stone sequence.

[†] Present address: Protein Pathways, 1145 Gayley Avenue, Ste 304, Los Angeles, California 90024, USA

As shown in Fig. 1, these links were combined with an additional 500 experimentally derived protein–protein interactions from the Database of Interacting Proteins (DIP)³ and the MIPS yeast genome database⁴, and 2,391 links among yeast proteins that catalyse sequential reactions in metabolic pathways⁵.

Of the total of 93,750 functional links found among 4,701 (76%) of the yeast proteins, we define 4,130 links to be of the ‘highest confidence’ (known to be correct by direct experimental techniques or validated by two of the three prediction techniques); 19,521 others are defined as ‘high confidence’ (predicted by phylogenetic profiles), and the remainder were predicted by either domain fusions or correlated mRNA expression, but not both.

We evaluated the quality of the links by recovery of known protein functions by prediction: we assume that if we link a protein, A, to a group of functionally related proteins, the shared functions of these other proteins provide a clue to the general function for A. Where the function of A is already known, we can test the predicted function. For this test, we chose the standardized keyword annotation of the Swiss-Prot database⁷ and systematically compared the known functions of all characterized yeast proteins with the function predicted by our methods. As one example chosen from the many yeast proteins tested, the Swiss-Prot keywords for the enzyme SAICAR synthetase (ADE1), which catalyses the seventh step of *de novo* purine biosynthesis, are ‘purine biosynthesis’ and ‘ligase’. Based upon the frequencies with which keywords appear in the annotation of the 18 proteins linked to ADE1, we predict the general function of ADE1 to be purine biosynthesis (13.6%), lyase (13.6%), transferase (11.4%) and ligase (6.8%). Therefore, we can use this analysis to predict the general biological process that a protein (such as ADE1) participates in, as well as to link the protein to many other

proteins of closely related function. We note that, for this example, none of the 18 proteins linked to ADE1 (which include ADE2, ADE5/7, ADE6, ADE8, ADE12, ADE13 and ADE16) shares any sequence similarity to ADE1, and only two pairs are similar to each other. Our results of systematic keyword analyses are listed in Table 1, along with confidence levels, data coverage and comparisons with random trials. The links verified by any two independent prediction techniques predict protein function with the same reliability as experimental interaction data and with over eight times the accuracy of random trials.

Functional links provide a means to characterize proteins of unknown function. There are 2,557 uncharacterized proteins in yeast⁴, proteins not studied experimentally and with no strong homologues of known function. Of these, 374 (or 15%) can be assigned a general function from the high and highest confidence functional links and 1,589 (or 62%) can be assigned a general function using all links.

A specific example of the assignment of function is shown in Fig. 2 for yeast open reading frame YGR021W, a member of a highly conserved protein family of unknown function. On the basis of the methods described here and the functional links they uncover to 28 other yeast proteins, this family can now be assigned a function related to mitochondrial protein synthesis. Two of the functional partners of YGR021W are also proteins in conserved families of unknown function: the *gidA* family and the *Caenorhabditis elegans* M02F4.4 family. These families, too, can now be associated with mitochondrial (or bacterial) protein synthesis. The link to triose-phosphate isomerase (Fig. 2) is of interest, considering the human myopathy in which a deficiency of this enzyme is correlated with grossly altered mitochondrial structure⁸.

Two additional examples of links are to the yeast prion Sup35 (ref. 9), and to MSH6, the yeast homologue of human colon-cancer related genes^{10,11}. In both cases, a general function is already known, but our method also predicts previously unknown functional links. In Fig. 3, the yeast prion Sup35, which acts as a translation release factor in its non-prion state, is linked with many proteins involved

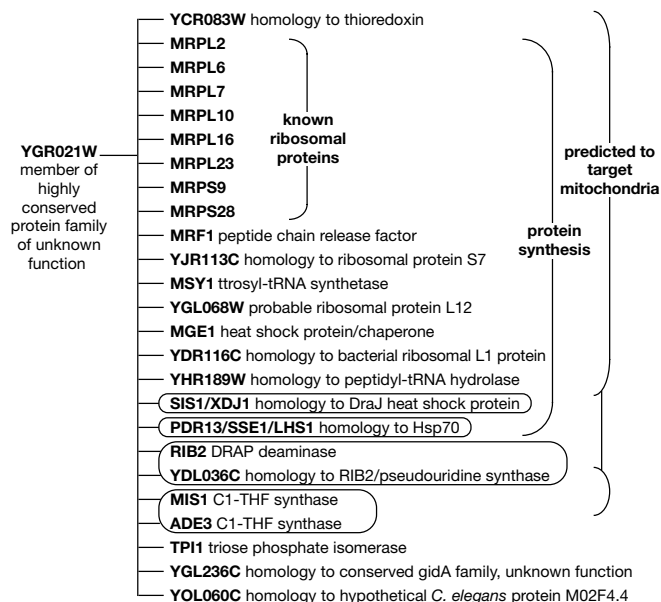


Figure 2 High confidence functional links found by phylogenetic profiles for the yeast protein YGR021W, a member of a protein family conserved in many organisms but of entirely unknown function. The links suggest that members of the YGR021W family operate in mitochondrial protein synthesis (brackets on the right). Of the 28 yeast proteins predicted to have a closely related function, 18 are also independently predicted to be mitochondrial²⁶, as is YGR021W, supporting the functional assignment. In Figs 2–4, only links from correlated evolution are given unless otherwise noted. Each protein is labelled by its genetic name (for example, MRPL7) or, if absent, by its yeast open reading frame code (for example, YGR021W), and a short description. Families of proteins with related sequences are circled. We note that the 28 linked proteins are all from yeast and are not homologues of YGR021W or (if not circled) of each other; instead, they are functionally linked.

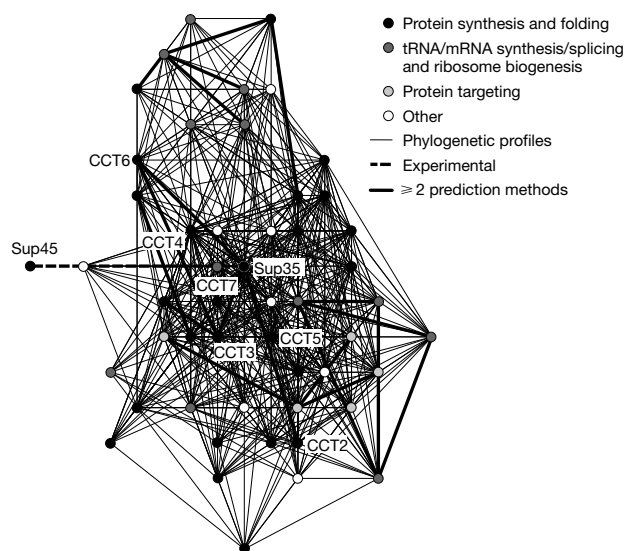


Figure 3 An illustration of the network of high (thin lines) and highest (bold lines) confidence links discovered among the proteins (circles) linked to the yeast prion and translation termination factor Sup35 (double circle). The functions linked to Sup35 are overwhelmingly related to ribosomes and protein synthesis, and include proteins with roles in protein folding, sorting, modification and targeting. For display purposes, proteins were modelled as points and links as springs, thereby positioning functionally related proteins close together.

Table 1 Reliability of functional assignments assessed by recovery of known protein function by prediction

	Number of proteins	Number of functional links	False positive rate* (%)	Ability to predict known function† (%)	Ability in random trials‡ (%)	Signal to noise ratio§
Individual prediction techniques						
Experimentall	484	500	6.5	33.2	4.0	8.3
Metabolic pathway neighbours	188	2,391	2.5	20.3	4.5	4.5
Phylogenetic profiles	1,976	20,749	29.5	33.1	7.4	4.5
Rosetta Stone method	1,898	45,502	36.4	26.5	7.7	3.4
Correlated mRNA expression	3,387	26,013	35.8	11.5	6.9	1.7
Combined predictions						
Links made by ≥2 prediction techniques	683	1,249	16.1	55.6	6.9	8.1
Highest confidence links	1,223	4,130	4.8	40.9	5.5	7.4
High confidence links	1,930	19,521	30.6	30.8	7.4	4.2
High and highest confidence links	2,356	23,651	21.8	32.0	6.8	4.7
All links	4,701	93,750	33.1	20.7	7.2	2.9

* The reliability of individual links was calculated as the percentage of pairwise links found between proteins of known function but having no functional categories in common (as tabulated in the MIPS database⁴, ignoring the functional categories 'unclassified' and 'classification not clear cut'). This estimate of false positives assumes complete knowledge of protein function and is therefore an upper limit. By this test, random links achieve a false positive rate of ~47%.

† The predictive power of individual techniques and combinations of techniques was evaluated by automated comparison of annotation keywords. By the methods listed, each protein is linked to one or more neighbour proteins. For characterized proteins ('query' proteins), the mean recovery of known Swiss-Prot keyword annotation by the keyword annotation of linked neighbours was calculated as:

$$(\text{keyword recovery}) = \frac{1}{A} \sum_{i=1}^A \sum_{j=1}^x \frac{n_j}{N} \quad (1)$$

where A is the number of annotated proteins, x is the number of query protein Swiss-Prot keywords, N is the total number of neighbour protein Swiss-Prot keywords, and n_j is the number of times query protein keyword j occurs in the neighbour protein annotation. Because functional annotations typically consist of multiple keywords, both specific and general, even truly related proteins show only a partial keyword overlap (for example, ~35%).

‡ Mean recovery of Swiss-Prot keyword annotation for query proteins of known function by Swiss-Prot keyword annotation of randomly chosen linked neighbours, calculated as in equation (1) for the same number of links as exist for real links (averages of 10 trials).

§ Calculated as ratio of known function recovered by real links to that recovered by random links. Although individual links have only moderate accuracy, combining information from many links significantly enhances prediction of function.

ll Experimentally observed yeast protein-protein interactions contained in the DIP³ and MIPS⁴ databases.

in protein synthesis, consistent with the primary role of Sup35, which interacts with the ribosome to release the newly synthesized peptide chain^{12,13}. Also linked to Sup35 are protein sorting and targeting proteins, consistent with an accessory role in guiding nascent proteins to their final cellular destinations. Sup35 shows both correlated evolution and correlated mRNA expression with the CCT chaperonin system, a yeast chaperonin system believed to aid folding of newly synthesized actin and microtubules¹⁴.

New links are also established for MSH6, a DNA mismatch repair protein¹⁵ whose human homologues, when mutated, cause most hereditary nonpolypoid colorectal cancers¹⁶. MSH6 is linked in Fig. 4 to the sequence-unrelated PMS1 DNA mismatch repair protein family, mutations of which are also tied to human colorectal cancer¹⁷. MSH6 is in turn linked via homologue MSH4 to the purine biosynthetic pathway by methylenetetrahydrofolate dehydrogenase¹⁸, to two RNA modification enzymes, and to an uncharacterized

protein family, which can now be investigated by taking nucleic acid repair or modification into consideration.

The methods of phylogenetic profiles¹ and domain fusions³, as well as gene localization^{19,20}, although they indirectly rely on sequence matching, discover links between non-homologous proteins, which provide much new information about functional relationships and hence go beyond the capabilities of traditional sequence matching. Analyses of mRNA co-expression, and potentially protein co-expression, are entirely independent of sequence matching and therefore provide information about proteins unique to the organism examined; this information is entirely inaccessible from sequence-matching techniques. The analysis of protein relationships by the methods we describe, especially when enhanced by other sources of functional relationships, should substantially speed the discovery of protein function. Functional predictions are available at <http://www.doe-mbi.ucla.edu>.

Methods

Experimental interactions

Pairwise links were created between yeast proteins, known from experimental literature to interact, using techniques such as co-immunoprecipitation and two-hybrid methods. Experimental observations include both direct *in vitro* measurements as well as indirect measurements of protein-protein interactions such as yeast two-hybrid data. We combined interaction data from the MIPS database⁴ and the DIP³, a community-developed database of protein-protein interactions. At the time this work was carried out, DIP contained 179 interactions between yeast proteins.

Linking of metabolic pathway neighbours

Yeast homologues of *Escherichia coli* proteins were found by BLAST²¹ homology searches. Pairwise links were defined between yeast proteins whose *E. coli* homologues catalyse sequential reactions (or one reaction step further away) in metabolic pathways, as defined in the EcoCyc database⁵.

Calculation of correlated evolution

Phylogenetic profiles were constructed for each yeast protein encoding the appearance of homologous amino-acid sequences in other organisms, as described in ref. 1, with some modifications. Instead of describing the presence or absence of yeast homologues in the full sequences genomes of 19 other organisms by a single bit, a real number expressing the evolutionary distance between the homologues (to be described elsewhere) was substituted, allowing functionally related proteins to be clustered by short euclidean distances between profile strings. The 20 completely sequenced genomes encoded in this way are available at the TIGR web site (<http://www.tigr.org/tdb/mdb/mdb.html>).

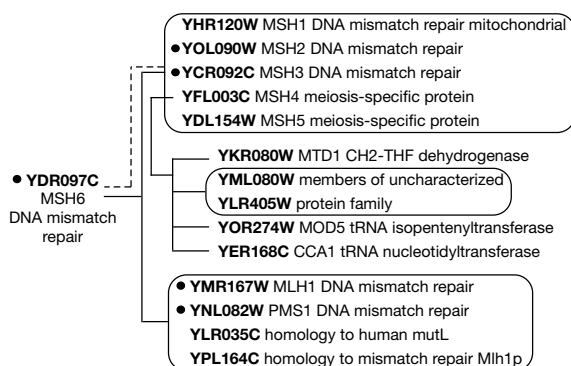


Figure 4 High and highest confidence functional links found for the yeast DNA repair protein MSH6, which is similar in sequence to colorectal cancer causing proteins in humans^{10,11}. Families of proteins with related sequences (circled), proteins implicated in human cancers (bullet points), and homology of MSH6 to the MSH DNA mismatch repair family (dashed line) are shown. MSH6 is also linked to MSH2 by correlated mRNA expression.

Calculation of correlated mRNA expression

Results of 97 individual publicly available DNA chip yeast mRNA expression data sets^{22–25} were encoded as a string of 97 numbers associated with each yeast open reading frame (ORF) describing how the mRNA of that ORF changed levels during normal growth, glucose starvation, sporulation and expression of mutant genes. This string is the analogue within one organism of a phylogenetic profile¹. The mRNA levels for each of the 97 experiments were normalized, and only genes that showed a two-standard-deviation change from the mean in at least one experiment were accepted, thereby ignoring genes that showed no change in expression levels for any experiment. Open reading frames with correlated expression patterns were grouped together by calculating the 97-dimensional euclidean distance that describes the similarity in mRNA expression patterns. Open reading frames were considered to be linked if they were among the 10 closest neighbours within a given distance cut-off, conditions that maximized the overlap of ORF annotation between neighbours.

Calculation of domain fusions

Proteins were linked by Rosetta Stone patterns as in ref. 3. Alignments were found with the program PSI-BLAST²¹. □

Received 7 May; accepted 23 August 1999.

- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96**, 4285–4288 (1999).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
- Marcotte, E. M. *et al.* Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
- Mewes, H. W., Hani, J., Pfeiffer, F. & Frishman, D. MIPS: a database for protein sequences and complete genomes. *Nucleic Acids Res.* **26**, 33–37 (1998).
- Karp, P., Riley, M., Paley, S. & Pellegrini-Toole, A. EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **26**, 50–53 (1998).
- The yeast genome directory. *Nature* **387** (suppl), 1–105 (1999).
- Bairoch, A. & Apewiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**, 49–54 (1999).
- Bardosi, A., Eber, S. W., Hendry, M. & Pekrun, A. Myopathy with altered mitochondria due to a triosephosphate isomerase (TPI) deficiency. *Acta Neuropathol. (Berl.)* **79**, 387–394 (1990).
- Wickner, R. B. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264**, 566–569 (1994).
- Miyaki, M. *et al.* Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature Struct. Biol.* **17**, 271–272 (1997).
- Fishel, R. *et al.* The human mutator gene homologue MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–1038 (1993).
- Kushirov, V. V. *et al.* Nucleotide sequence of the Sup2(Sup35) gene of *Saccharomyces cerevisiae*. *Gene* **66**, 45–54 (1988).
- Stansfield, I. *et al.* The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4365–4373 (1995).
- Chen, X., Sullivan, D. S. & Huffaker, T. C. Two yeast genes with similarity to TCP-1 are required for microtubule and actin function *in vivo*. *Proc. Natl Acad. Sci. USA* **91**, 9111–9115 (1994).
- Johnson, R. E., Kovvali, G. K., Prakash, L. & Prakash, S. Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability. *J. Biol. Chem.* **271**, 7285–7288 (1996).
- Lynch, H. T., Fusaro, R. M. & Lynch, J. F. Cancer genetics in the new era of molecular biology. *Ann. NY Acad. Sci.* **833**, 1–28 (1997).
- Papadopoulos, N. *et al.* Mutations of a MutL homolog in hereditary colon cancer. *Science* **263**, 1625–1629 (1994).
- West, M. G., Horne, D. W. & Appling, D. R. Metabolic role of cytoplasmic isozymes of 5,10-methylenetetrahydrofolate dehydrogenase in *Saccharomyces cerevisiae*. *Biochemistry* **35**, 3122–3132 (1996).
- Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. *Proc. Natl Acad. Sci. USA* **96**, 2896–2901 (1999).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1998).
- Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
- Myers, L. C., Gustafsson, C. M., Hayashibara, K. C., Brown, P. O. & Kornberg, R. D. Mediator protein mutations that selectively abolish activated transcription. *Proc. Natl Acad. Sci. USA* **96**, 67–72 (1999).
- Horton, P. & Nakai, K. Better prediction of protein cellular localization sites with the *k* nearest neighbors classifier. *Intell. Sys. Molec. Biol.* **5**, 147–152 (1997).

Acknowledgements

This work was supported by a Department of Energy/Oak Ridge Institute for Science and Education Hollaender postdoctoral Fellowship (E.M.), a Sloan Foundation/Department of Energy postdoctoral fellowship (M.P.), and grants from the DOE.

Correspondence and requests for materials should be addressed to D.E. (e-mail: david@mbi.ucla.edu).

Protein interaction maps for complete genomes based on gene fusion events

Anton J. Enright, Ioannis Iliopoulos, Nikos C. Kyrpides* & Christos A. Ouzounis

Computational Genomics Group, Research Programme, The European Bioinformatics Institute, EMBL Cambridge Outstation, Cambridge CB10 1SD, UK

* Integrated Genomics Inc., 2201 West Campbell Park Drive, Chicago, Illinois 60612, USA

A large-scale effort to measure, detect and analyse protein–protein interactions using experimental methods is under way^{1,2}. These include biochemistry such as co-immunoprecipitation or crosslinking, molecular biology such as the two-hybrid system or phage display, and genetics such as unlinked noncomplementing mutant detection³. Using the two-hybrid system⁴, an international effort to analyse the complete yeast genome is in progress⁵. Evidently, all these approaches are tedious, labour intensive and inaccurate⁶. From a computational perspective, the question is how can we predict that two proteins interact from structure or sequence alone. Here we present a method that identifies gene-fusion events in complete genomes, solely based on sequence comparison. Because there must be selective pressure for certain genes to be fused over the course of evolution, we are able to predict functional associations of proteins. We show that 215 genes or proteins in the complete genomes of *Escherichia coli*,

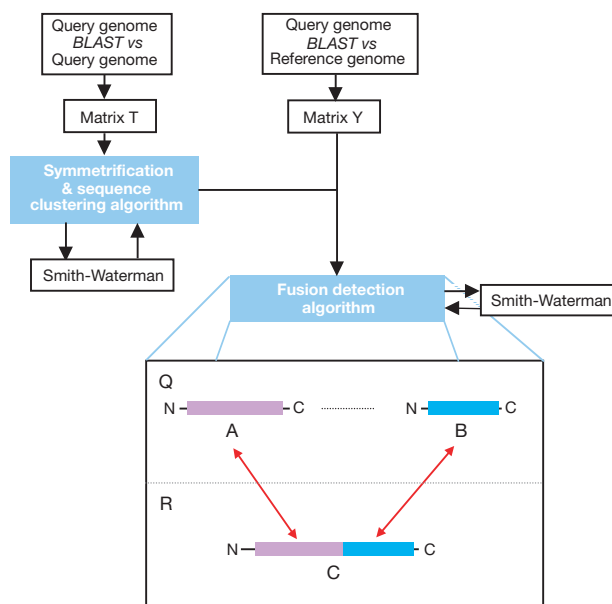


Figure 1 Flowchart of the algorithm. All similarities within the query genome Q detected using BLAST²³ are stored in matrix T. For all nonsymmetrical hits, an additional Smith–Waterman comparison²⁶ is used to resolve false negatives. The query genome is then compared with the reference genome (or database), and similarities are stored in matrix Y. The fusion-detection algorithm identifies cases of the form depicted in the inset, where query (Q) proteins A and B exhibit similarity to reference (R) protein C by checking matrix Y, but not to each other, by checking matrix T (which is further confirmed by an additional Smith–Waterman comparison). Both Smith–Waterman runs are executed an additional 25 times, with randomization of the sequences, and a Z-score is obtained: if the Z-score is higher than a certain threshold, the similarity is accepted as significant. The ‘key’ abstraction is that a candidate pair (A,B) of query proteins can either represent a false-negative, or a component pair matching the composite protein C. Total computation time is ~4 h on an SGI Octane two-processor workstation.

***Haemophilus influenzae* and *Methanococcus jannaschii* are involved in 64 unique fusion events. The approach is general, and can be applied even to genes of unknown function.**

Previously, the only computational approach to the problem of protein-protein interactions has been the study of subunit interfaces, using information from the protein structure database⁷⁻⁹. Recently, a number of studies have exploited the ordering of related

genes in genome sequences as a basis for the identification of interacting proteins^{10,11}. This approach relies on the notion that gene proximity is a result of selective pressure to associate genes that are co-regulated and possibly interacting^{10,12}. These analyses result in a number of false predictions, however, because the constraint of proximity is not strong, and interactions between products of distantly located genes are not identifiable. In addition, this

Table 1 The 64 fusion events in the genomes of *E. coli*, *H. influenzae* and *M. jannaschii* detected on the basis of composite proteins in these three genomes and the genome of *S. cerevisiae*.

Case	Component	Component	Composite	EC	HI	MJ	SC	N
1	GalE	GalM	GAL10	▲▼	▲▼		●1	2
2	AccC	B0712-hypothetical	DUR1,2	▲▼	▲▼		●1	2
3	Hypothetical	Hypothetical	PYC2, PYC1			▲▼	●2	1
4	HisH	HisF	HIS7	▲▼	▲▼	▲▼	●1	3
5	His(E)	HisD	HIS4	▲▼	▲▼	▲▼	●1	3
6	RpoA'	RpoA''	RPO21, RPO31, RPA190			▲▼	●3	1
7	GltB	GltD	GLT1	▲▼			●1	1
8	AroB/AroA/AroK/AroD/AroE	Multiple fusion	ARO1	▲▼▲▼▲	▲▼▲□□	□▼□□▲	●1	3
9	Aconitase subunit	Aconitase subunit	LYS4			▲▼	●1	1
10	ArgA	ArgC	ARG5,6	▲▼			●1	1
11	LeuC	LeuD	LEU1	▲▼	▲▼	▲▼	●1	3
12	TrpA	TrpB	TRP5	▲▼	▲▼	▲▼	●1	3
13	PurD	PurM	ADE5,7	▲▼	▲▼	▲▼	●1	3
14	PurL	PurQ	ADE6	●1	●1	▲▼	●1	1
15	CarA/CarB/PyrB	Multiple fusion	URA2	▲▼▲		▲▼▲	●1	2
16	B1378	CysI	ECM17	▲▼			●1	1
17	TrpG	TrpC	TRP3	▲▼	▲▼	▲▼	●1	3
18	AgaG	AgaF	HyuA, HuyB			▲▼	●1	1
19	IlvG_1	IlvG_2	ILV2	▲▼	●1	●2	●1	1
20	GmpA	GmpB	GUA1	●1	●1	▲▼	●1	1
21	GyrB, ParE	GyrA, ParC	TOP2	▲▼▲▼	▲▼▲▼		●1	4
22	FolK	FolP	Folate biosynthesis (probable)	▲▼	▲▼		●1	3
23	PabA	PabB	ABZ1	▲▼	▲▼		●1	2
24	PurK	PurE	ADE2	▲▼	▲▼	▲▼	●1	3
25	RpoB'	RpoB'	RPB2, RET1, A135	●1	●1	▲▼	●3	1
26	ThiE	ThiM	THI6		▲▼		●1	1
27	ThiD	TenA	thi21, thi20, thi22		▲▼		●3	1
28	TkIA	TkIB	TKL1, TKL2	●2	●1	▲▼	●2	1
29	LysC	hom	ThrA, MetL	●2	●1		▲▼	1
30	ABC transporter	Hypothetical	B0879	●1		▲▼▲▼		4
31	Hypothetical	Putative methyltransferase	B0948	●1		▲▼▲▼		2
32	TrpD	TrpD	TrpD	●1	▲▼	▲▼		2
33	FumA	FumB	FumA	●1		▲▼		1
34	Hypothetical tkt	PheA	PheA	●1	●1	▲▼		1
35	FprA	Rubredoxin	B2710	●1		▲▼▲▼		6
36	TrxM	Hypothetical	B0492	●1	▲▼			1
37	Hypothetical	Hypothetical	B1816, B2063	●2	▲▼▲▼			3
38	CpxR, YgiX	TyrR	AtoC, YfhA, GlnG, HydG	●4	▲▼▲▼			2
39	Hypothetical	Multiple fusion	B2324	●1	▲▼▲			1
40	Hypothetical	Hypothetical	B2474	●1	▲▼			1
41	Hypothetical	Hypothetical	SufI	●1	▲▼			1
42	HemX	Hypothetical	HemX	●1	▲▼			1
43	TrpC	TrpF	TrpC	●1	●1	▲▼		1
44	CitX	CitG	CitG	▲▼	●1			1
45	SbmA	Hypothetical	ABC transporter/ATP-binding	▲▼▲▼▲▼▲	●1			7
46	B3777	B3776	Hypothetical	▲▼	●1			1
47	B2612	YfjD	Hypothetical	▲▼	●1			1
48	YgfQ	YgfR	Hypothetical	▲▼	●1	●1		1
49	YabK	B0263	Hypothetical	▲▼	●1			1
50	YhaQ	YhaP	SdaA	▲▼	●1			1
51	YbfH	YbfG	Hypothetical	▲▼	●1			1
52	PurF	YhfN	GlmS	▲▼	●1			1
53	FrwB, FrwD	FrwC, B2386	FruA	▲▼▲▼	●1			4
54	UgpC	YtfS	RbsA, MglA	▲▼	●2			1
55	B1515, B1899	AraH	RbsC, MglC	▲▼	●2			2
56	NrfF	NrfG	NrfF	▲▼	●1			1
57	MsrA	B1778	MsrA	▲▼	●1			1
58	SbmA	Hypothetical	ABC transporter/ATP-binding	▲▼▲▼▲▼▲	●1			7
59	YhgK	YhgJ	Probable RNA cyclase	▲▼		●1		1
60	FrdB	GlpC	Iron-sulfur-binding reductase	▲▼		●1		1
61	RffH, RfbA, GalF, GalU	B0359	Glucose-1-P thymidyltransferase	▲▼▲▼		●1		4
62	B3016	B3015	Hypothetical	▲▼		●1		1
63	LeuS	YgiH	MetS	▲▼		●1		1
64	TopB	YrdD	TopA	▲▼	▲▼	●1		2

122

The component gene/protein names (or identifiers) and the composite (fusion) gene/protein names (or identifiers) are listed. Columns EC, HI, MJ and SC correspond to *E. coli*, *H. influenzae*, *M. jannaschii* and *S. cerevisiae*, respectively; N lists the maximum number of possible pairwise interactions taking into account paralogy in the query genomes (multiple-component cases are counted as a single interaction). Symbols represent a corresponding component or composite genes/proteins: triangle pairs, ▲▼, a pair of component proteins in the query genome predicted to interact based on their similarity to a composite protein in the reference genome; alternating triangles, ▲▼▲▼▲▼▲, multiple-component genes/proteins (cases 8, 15 and 39); open squares, □, absence of a component from a multiple-fusion event (case 8); consecutive triangles, ▲▲▲/▼▼▼, the exact number of detected paralogous component genes/proteins in the query genome; filled circles, composite genes/proteins, the number represents the number of paralogous composite genes/proteins in the reference genome. The sort order follows the three species against the composite-protein sequence identifiers for the yeast genome, and then the other three species in succession. Genes are named where possible; where none is available, the sequence identifier is used instead. All fusions were confirmed by reverse BLAST searches using the composite protein as query, which identified all the component proteins. Note that functional annotation is not necessary but frequently useful in resolving paralogous cases (for example, case 21). Predictions imply functional associations and not necessarily direct molecular interactions. For gene/protein identifiers and references for the known cases, see Supplementary Information.

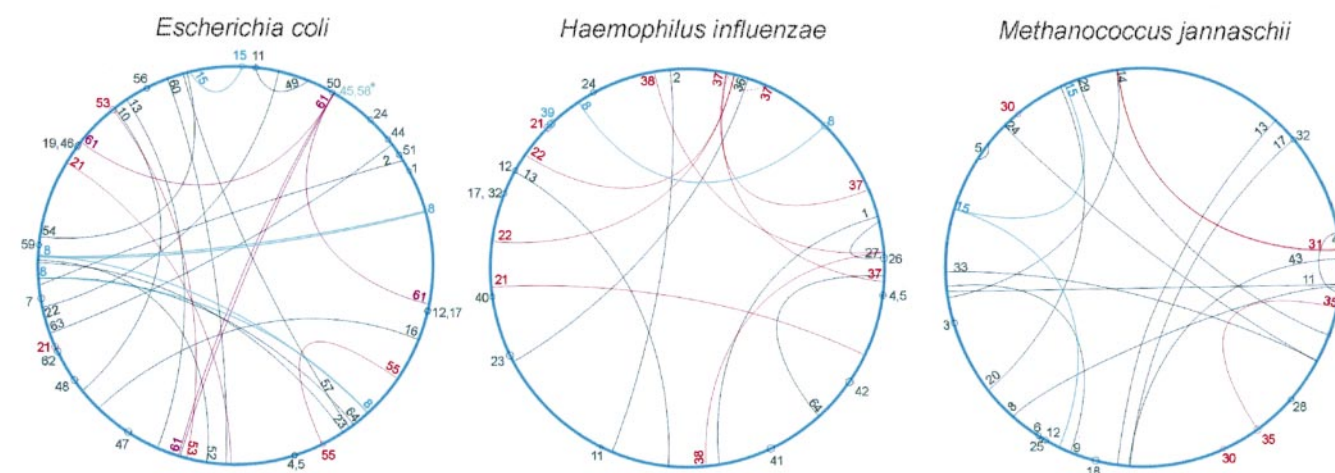


Figure 2 Representation of protein interaction maps for the most likely interactions predicted for *E. coli*, *H. influenzae* and *M. jannaschii*. In the large blue circles, which represent the three genomes, 0° corresponds to the first base pair, and 360° the last base pair of the genome. Predicted interactions are indicated by linking the circular map positions of the genes involved. In cases of neighbouring genes (<5°), a small circle indicates the predicted interaction between two genes at that region; otherwise, an arc links the two genes in question. Multiple interactions are not cross-labelled. Some

paralogous cases are resolved and only the most likely case is indicated by an arc. All cases are numbered according to Table 1. Predictions are colour coded: black, pairwise interactions; blue, multiple interactions; red/purple, cases where, due to paralogy, more than one pairwise interaction is possible (red, two possibilities; purple, more than two possibilities); green (marked by asterisk), because of a large number of paralogs, no interaction can be easily resolved. The source of the prediction (composite protein from a given species) is not indicated.

approach may not be applicable to eukaryotes, because the co-regulation of genes is not imposed at the genome structure level. Another approach, based on the mere presence/absence of genes in different species, also attempts to identify functionally associated genes¹³.

Our method overcomes some of these problems by exploiting the fact that certain protein families in a given species consist of fused domains that usually correspond to single, full-length proteins in other species. We define these proteins as 'composite' proteins (otherwise known as fusion proteins) and their individual domains as 'component' proteins. We rely primarily on complete genome sequences for the identification of fusion events because we can detect 'orthologous' (equivalent) proteins across species. The underlying assumption is that if a composite protein is uniquely similar to two component proteins in another species, which may not necessarily be encoded by neighbouring genes, the component proteins are most likely to interact (Fig. 1; see Methods). We use the term 'interaction' to imply either direct physical interaction or an indirect functional association (for example, involvement in the same biochemical pathway or similar gene regulation).

The situation is comparable to experimental approaches that make use of artificial constructs of fused genes for biochemical analysis and protein-purification technology^{14,15}. The process has also been observed in evolution, perhaps the most widely known example being the fusion of tryptophan synthetase α - and β -subunits from bacteria to fungi¹⁶.

We identified 215 component proteins in the 3 query genomes that are involved in 88 fusion events (of which 64 are unique) (Fig. 2; triangles in Table 1). There are 39 fusion events in *E. coli*, 24 in *H. influenzae* and 25 in *M. jannaschii*, with 2.44 proteins per fusion event on average (due to both multiple-fusion events, for example, case 8, and paralogous genes, for example, case 21, Table 1).

We identified 94 composite proteins in the 4 reference genomes (representing 77 fusion cases or protein families, circles in Table 1; there are 17 paralogous composite proteins). The *E. coli* genome contains only 24 composite proteins (18 fusion-protein families) with reference to the component proteins, as opposed to the much smaller *H. influenzae* genome which contains 25 composite proteins (23 fusion-protein families). In contrast, the *M. jannaschii* genome has only 9 composite proteins (8 fusion-protein families). The

remaining 36 composite proteins (28 fusion-protein families) are detected in the genome of *Saccharomyces cerevisiae*. Because query genomes were in turn used as reference genomes (see Methods), we detected eight cases shared between them: *E. coli* and *H. influenzae* share six; *E. coli* and *M. jannaschii* none; and *H. influenzae* and *M. jannaschii* two (Table 1).

There are only three multiple-fusion events (case 8 deriving from the yeast ARO1 gene¹⁷, case 15 deriving from the yeast URA2 gene¹⁸ and case 39 based on the *E. coli* gene 2282 of unknown function) (Table 1). The total number of possible pairwise interactions can be as many as 122 (column N in Table 1), depending on the degree of paralogy for certain proteins, which introduces some uncertainty in the prediction. Paralogy in the component proteins increases the number of possible interactions (for example, case 45), thereby decreasing the certainty of the prediction. Conversely, paralogy in the composite proteins increases the certainty that the component proteins interact (for example, case 27 (ref. 19)), because the fusion event is repeatedly observed within (or even across) genomes.

Notably, as many predicted interactions occur between products of distant genes as between products of neighbouring genes (Fig. 2): there are only 8 interactions between products of neighbouring genes in *M. jannaschii*, 14 in *H. influenzae* and 18 in *E. coli* (Fig. 2). This underlines the potential of our method to identify interacting proteins in complete genome sequences beyond the simple proximity constraint^{10,11}. Some of the interactions between products of neighbouring genes are cases of well-known gene pairs (for example, case 6, the DNA-dependent RNA polymerase σ / σ' pair in *M. jannaschii*), although there are also cases that may be sequencing artefacts (for example, case 19 in Table 1, which is 'intact' in all species except *E. coli*).

The predicted interactions display some complex patterns of distribution along the genome sequences (Fig. 2). For example, certain regions seem to contain a disproportionate number of genes or proteins involved in fusion events (for example, cases 12, 17 and 32 representing the tryptophan biosynthesis gene cluster). Also, some intricate symmetries occur, deriving from multiple-fusion events or paralogous proteins (for example, cases 8 and 61 in *E. coli*, respectively).

Our method identifies a number of well-known interacting partners (for example, cases 4–13). In total, 26 out of the 64 cases

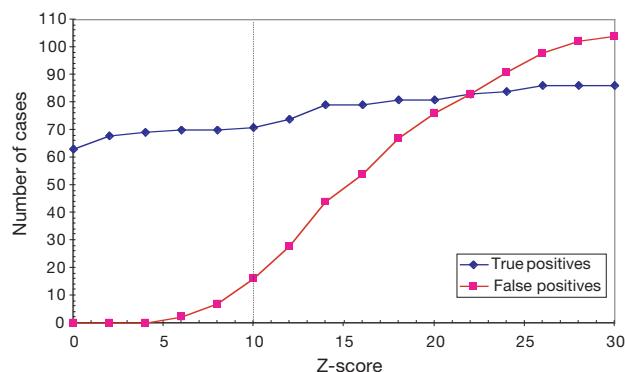


Figure 3 The dependence of the number of true and false positive hits with respect to the Z-score threshold used. Dashed line indicates the threshold used. True positives are 'unrelated' genes that are involved in a fusion event and are candidates for pairwise interactions; false positives are 'related' genes that are also accepted. The decision of whether two genes are related or not (in terms of sequence similarity) depends on the Z-score value: the higher that value is, the more permissive the criterion becomes for two related genes to be considered. True positives are all cases listed in Table 1. False positives were eliminated manually by inspection of the alignment overlap between the two component genes. For Z-scores <10, all cases have been manually and automatically verified. For Z-scores >10, all cases have been only automatically checked; therefore, these values represent an upper bound for true positives.

(40%) listed in Table 1 are involved in the same protein complex or biochemical process (for references, see Supplementary Information). A number of unconfirmed cases (for example, cases 31, 57 and 60, Table 1) constitute some interesting, testable predictions; for example, case 60 represents a predicted interaction between gene products FrdB (fumarate reductase) and GlpC (heterodisulfide reductase) in *E. coli*. We obtained 85 fusion events of which 64 are unique and 21 are false positives (Fig. 3); thus, the precision of the method can be estimated as 75% (64/85) with the current parameter settings (see Methods).

Coverage cannot easily be estimated, as we do not know in advance how many proteins potentially interact within the query genome. We also lack a standard set of fusion events to estimate coverage for the given query genomes. We have tried to estimate coverage by counting the number of false negatives (11, Table 1; see Supplementary Information). Thus, the maximum estimate for the coverage is as high as 95% (215/226) with the current parameter settings and the above assumption. Another false-negative case, fatty acid synthase (and its bacterial homologues)²⁰, is caused by low sequence similarity relationships. Precision and coverage can be controlled by modifying the cut-off scores in the post-processing of the homology searches (see Methods).

From a total of 7,768 protein sequences in the 3 complete query genomes, the minimum number of components involved in a multidomain-fusion event is 215, or 2.8%. Most of these proteins of known function appear to be metabolic enzymes, an effect possibly due to metabolic channelling of substrates²¹. Our results, together with another study²², provide the first estimates for the numbers of gene-fusion events based on complete-genome comparisons. By adding more reference genomes, or using the complete database, this number is expected to increase. We are in the process of repeating this analysis to identify fusion events and possible interactions for the genomes of *S. cerevisiae* and *Caenorhabditis elegans*. With sufficient computational power, this analysis could be performed for all complete genomes. The exactness and high efficiency of the method makes it applicable to proteomics research, and complementary to continuing experimental approaches for the identification of protein interactions. We believe that this general approach, if used wisely, will find a number of applications in genomics and computational biology. □

Methods

Genome sequences

The genome sequences (and total number of open reading frames (ORFs)) used were *E. coli* (4290; ftp://genetics.wisc.edu/pub/sequence/m52p.fap.gz); *H. influenzae* (1707; ftp://ftp.tigr.org/pub/data/h_influenzae/GHI.pep.gz); *M. jannaschii* (1771; ftp://ftp.tigr.org/pub/data/m_jannaschii/GMJ.pep.gz); and *S. cerevisiae* (6243; ftp://genome-ftp.stanford.edu/pub/yeast/yeast_ORFs/orf_trans.fasta.Z).

Genome comparison

The most reliable prediction of protein-protein interactions is that within complete genomes, and we only considered query databases that have this property. In other words, we did not address the general problem of protein interaction predictions within any arbitrary database. Both query and reference databases are always considered to be complete-genome databases. Generalizations of this formalism are possible and query and reference databases may be interchangeable. The advantage of performing this analysis only within complete genomes is that these data are both comprehensive (no better candidate may be identified) and unbiased (no cases occur in which more fusions will be detected because of experimental sampling). Each of the three complete genomes of *E. coli*, *H. influenzae* and *M. jannaschii* were used in turn as the query genome Q; Q was compared with the other two genomes, plus the genome of the yeast *S. cerevisiae*, as reference genomes.

Sequence analysis

The query database is compared against itself using BLASTP (v 2.0)²³, after masking compositionally biased regions (using the CAST algorithm²⁴; and V. Promponas *et al.*, manuscript in preparation), and all pairwise sequence similarities are recorded in a binary matrix T (Fig. 1). The matrix is symmetrified (similarly to that in ref. 25), using a Smith-Waterman²⁶ dynamic-programming alignment algorithm²⁷, which is executed only for nonsymmetrical pairs. The query database is also compared against the reference database, as above, and similarities are recorded in binary matrix Y (Fig. 1). For all entries C in the reference database, entry pairs (A,B) from the query database similar to reference entry C are collected (Fig. 1). Every pair (A,B) similar to C is looked up in the self-comparison matrix T; if dissimilar, it is further checked for similarity using a second dynamic-programming alignment pass to eliminate the possibility that it was a false-negative case during the initial self-comparison phase. If not (see Fig. 1 for details), then A and B in the query database (usually a complete genome) are candidates for a fusion event and can be predicted to interact. It is apparent that, although the coverage of the query database may strongly depend on the reference database, the precision can be very high. Unfortunately, no pertinent data set exists to estimate precision. Precision and coverage are controlled by a Z-score parameter setting (Fig. 3).

Received 15 July; accepted 15 September 1999.

- Mendelsohn, A. R. & Brent, R. Protein interaction methods—toward an endgame. *Science* **284**, 1948–1950 (1999).
- Blackstock, W. P. & Weir, M. P. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**, 121–127 (1999).
- Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123 (1995).
- Luban, J. & Goff, S. P. The yeast two-hybrid system for studying protein-protein interactions. *Curr. Opin. Biotechnol.* **6**, 59–64 (1995).
- Lecrenier, N., Foury, F. & Goffeau, A. Two-hybrid systematic screening of the yeast proteome. *BioEssays* **20**, 1–5 (1998).
- Lakey, J. H. & Raggett, E. M. Measuring protein-protein interactions. *Curr. Opin. Struct. Biol.* **8**, 119–123 (1998).
- Janin, J. & Rodier, F. Protein-protein interaction at crystal contacts. *Proteins* **23**, 580–587 (1995).
- Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
- Larsen, T. A., Olson, A. J. & Goodsell, D. S. Morphology of protein-protein interfaces. *Structure* **6**, 421–427 (1998).
- Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a finger-print of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. *Proc. Natl Acad. Sci. USA* **96**, 2896–2901 (1999).
- Tamames, J., Casari, G., Ouzounis, C. & Valencia, A. Conserved clusters of functionally related genes in two bacterial genomes. *J. Mol. Evol.* **44**, 66–73 (1997).
- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96**, 4285–4288 (1999).
- Bülow, I. Preparation of artificial bifunctional enzymes by gene fusion. *Biochem. Soc. Symp.* **57**, 123–133 (1990).
- Wales, M. E. & Wild, J. R. Analysis of structure-function relationships by formation of chimeric enzymes produced by gene fusion. *Methods Enzymol.* **202**, 687–706 (1991).
- Burns, D. M., Horn, V., Paluh, J. & Yanofsky, C. Evolution of the tryptophan synthetase of fungi. Analysis of experimentally fused *Escherichia coli* tryptophan synthetase alpha and beta chains. *J. Biol. Chem.* **265**, 2060–2069 (1990).
- Duncan, K., Edwards, R. M. & Coggins, J. R. The *Saccharomyces cerevisiae* ARO1 gene. An example of the co-ordinate regulation of five enzymes on a single biosynthetic pathway. *FEBS Lett.* **241**, 83–88 (1988).
- Denis-Duphil, M. Pyrimidine biosynthesis in *Saccharomyces cerevisiae*: the *ura2* cluster gene, its multifunctional enzyme product, and other structural or regulatory genes involved in *de novo* UMP synthesis. *Biochem. Cell Biol.* **67**, 612–631 (1989).
- Ouzounis, C. A. & Kyriakides, N. C. ThiD-TenA: a gene pair fusion in eukaryotes. *J. Mol. Evol.* **45**, 708–711 (1997).

20. Smith, S. The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J.* **8**, 1248–1259 (1994).
21. Welch, G. R. & Easterby, J. S. Metabolic channeling versus free diffusion: transition-time analysis. *Trends Biochem. Sci.* **19**, 193–197 (1994).
22. Marcotte, E. M. *et al.* Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Andrade, M. A. *et al.* Automated genome sequence analysis and annotation. *Bioinformatics* **15**, 391–412 (1999).
25. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
26. Smith, T. F. & Waterman, M. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
27. Pearson, W. R. Effective protein sequence comparison. *Methods Enzymol.* **266**, 227–258 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Carroll and S. Searle for technical advice. This work was supported by the European Molecular Biology Laboratory and the TMR Programme of the European Commission DGXII (Science, Research and Development). Patent application filed on behalf of the European Molecular Biology Laboratory.

Correspondence and requests for materials should be addressed to C.A.O. (e-mail: ouzounis@ebi.ac.uk).

Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells

Maurizio Molinari & Ari Helenius

Swiss Federal Institute of Technology Zurich, Institute of Biochemistry, Universitätsstrasse 16, CH-8092 Zurich, Switzerland

The formation of intra- and interchain disulphide bonds constitutes an integral part of the maturation of most secretory and membrane-bound proteins in the endoplasmic reticulum^{1,2}. Evidence indicates that members of the protein disulphide isomerase (PDI) superfamily are part of the machinery needed for proper oxidation and isomerization of disulphide bonds^{3–6}. Models based on *in vitro* studies predict that the formation of mixed disulphide bonds between oxidoreductase and substrate is intermediate in the generation of the native intrachain disulphide bond in the substrate polypeptide⁷. Whether this is how thiol oxidoreductases work inside the endoplasmic reticulum is not clear. Nor has it been established which of the many members of the PDI superfamily interacts directly with newly synthesized substrate proteins, because transient mixed disulphides have never been observed in the mammalian endoplasmic reticulum during oxidative protein folding^{7,8}. Here we describe the mechanisms involved in co- and post-translational protein oxidation *in vivo*. We show that the endoplasmic-reticulum-resident oxidoreductases PDI and Erp57 are directly involved in disulphide oxidation and isomerization, and, together with the lectins calnexin and calreticulin, are central in glycoprotein folding in the endoplasmic reticulum of mammalian cells.

To determine whether maturation of proteins in the endoplasmic reticulum (ER) of live cells is accompanied by formation of mixed disulphides containing ER oxidoreductases, the folding of viral glycoproteins E1 and p62 expressed in Semliki Forest virus (SFV)-infected mammalian cells was monitored. Four hours after infection, Chinese hamster ovary (CHO) cells were pulse-labelled for 2 min with ³⁵S-methionine and -cysteine (0.5 mCi per dish). After

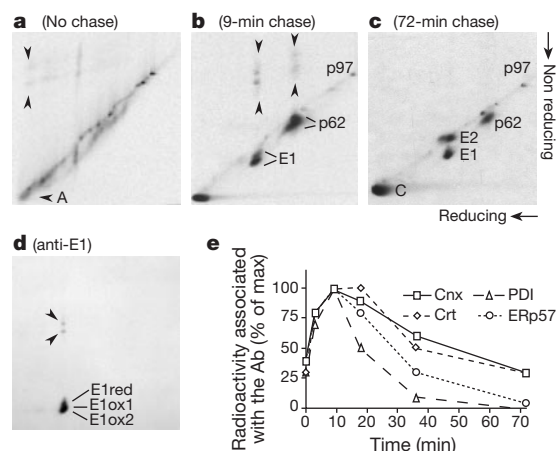


Figure 1 Cotranslational folding of SFV glycoproteins analysed by 2D SDS–PAGE. **a**, Viral growing nascent chains are visible on, above and below the gel diagonal. Mixed disulphides and intramolecular disulphide bonds formed cotranslationally (arrowheads indicate the shortest viral nascent chains engaged in intermolecular disulphide-bonded complexes; arrow A indicates the shortest viral nascent chains acquiring intramolecular disulphide bonds). **b**, **c**, Upon chase, viral proteins are terminated and radioactivity condenses in full-length viral proteins. **d**, Postnuclear supernatants from **b** were precipitated with a monoclonal antibody to E1. E1 in three different oxidation states (E1red, E1ox1 and E1ox2), as well as E1-containing mixed disulphides (arrowheads), are indicated. **e**, Kinetics of viral glycoprotein binding/release from Cnx, Crt, PDI and Erp57. Ab, antibody.

various chase times, cells were flooded with a membrane-permeable alkylating agent, *N*-ethylmaleimide (NEM, 20 mM in PBS, pH 6.8), to prevent post-lysis oxidation of free cysteines, and to trap mixed disulphides. NEM has been extensively used as a rapidly acting, *in vivo* alkylating agent of cysteines⁹, and its efficiency has been confirmed¹⁰. Similar results were obtained when we solubilized pulse-labelled cells at acidic pH (detergent and 20% formic acid) as an alternative method to block cysteine reactivity. Under these conditions, the SH form, which predominates at low pH, is not capable of attacking existing disulphides¹¹. Postnuclear supernatants (PNS) were prepared with nonionic detergent essentially as described⁹, and analysed using a sensitive two-dimensional (2D) gel electrophoresis technique, in which the first dimension was run under nonreducing conditions in a 1.5-mm diameter capillary tube. For the second dimension, the gel extruded from the glass capillary was boiled for 10 min in reducing sample buffer and placed on a standard 8% SDS polyacrylamide gel electrophoresis (PAGE) gel^{12,13}.

The 2D gels shown in Fig. 1a–c indicate the progression of synthesis, disulphide bond formation and processing of SFV proteins. They allow identification of intermolecular disulphide-bonded complexes, which run above the gel diagonal, and proteins with intrachain disulphides which run below. The simplest pattern is seen in Fig. 1c, after 72 min of chase. Five major radioactive spots are visible. All of these are viral proteins, because SFV induces a complete block in host protein synthesis¹⁴. One corresponds to the capsid protein (C) and one to p97, a nontranslocated side product of viral glycoprotein synthesis¹⁵. Capsid and p97 are cytosolic, and their positions on the gel diagonal indicate that they do not contain intrachain disulphide bonds. The three other spots, located below the diagonal, correspond to type I membrane glycoproteins E1 (16 cysteine residues, 1 N-linked glycan), p62 (22 cysteine residues, 4 N-linked glycans) and E2, a cleavage product of p62 formed in the Golgi compartment. E1 and p62 are the only proteins in the infected cells that are translocated into the ER and processed by the ER folding machinery.

After 72 min of chase, folding of SFV glycoproteins was completed and E1, p62 and E2 had acquired a conformation that made

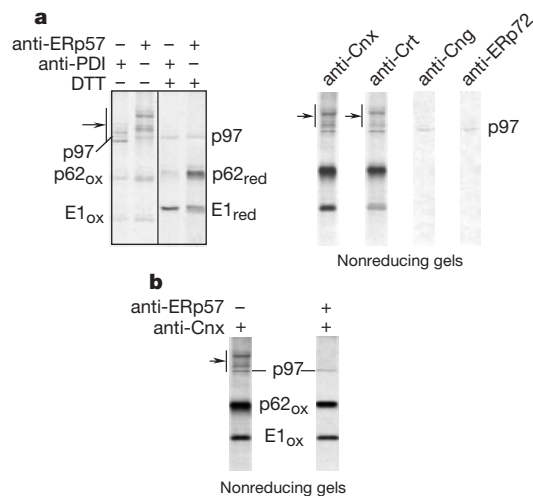


Figure 2 PDI and Erp57 form mixed disulphides and have different substrate specificity. **a**, Postnuclear supernatants (2 min pulse, 9 min chase) were precipitated with antibodies to PDI (left panel, lanes 1 and 3) and Erp57 (left panel, lanes 2 and 4). PDI- (lane 1) and Erp57 (lane 2)-containing complexes were reduced (lanes 3 and 4, respectively), which established that E1 was the major labelled component associated with PDI in mixed disulphides; both E1 and p62 were associated with Erp57. Disulphide-bonded complexes were also precipitated with Cnx and Crt antibodies, whereas no detectable labelled viral protein was associated with colligin-Hsp47 and Erp72 (right panel). **b**, A double immunoprecipitation (anti-Erp57 followed by anti-Cnx) showed that the disulphide-bonded complexes that precipitated with Cnx were associated with Erp57. Arrows indicate disulphide-bonded complexes.

them resistant to 5 mM dithiothreitol (DTT)⁹ (data not shown). As no radioactivity was visible above the diagonal, none of the labelled viral proteins formed interchain disulphide-bonded complexes at this time of chase.

After 9 min of chase (Fig. 1b), much earlier in the maturation process, the absence of E2 indicated that the glycoproteins were still in the ER. E1 and p62 were present in different oxidation states, as indicated by heterogeneous mobility below the gel diagonal. The most striking difference to the 72-min chase time was the clustered spots of radioactivity above the diagonal of the 2D gel, precisely above E1 and p62 (arrowheads). This showed that a small fraction of the glycoproteins were in intermolecular disulphide-bonded complexes. Notably, both of the glycoproteins in the ER were engaged in complexes, whereas the capsid protein, which is located in the cytosol, was not.

Immunoprecipitation with specific E1 antibodies brought down incompletely folded E1 monomers in different oxidation states (E1_{red}, E1_{ox1} and E1_{ox2}, Fig. 1d) and the complexes above E1 (arrowheads). This indicated that E1 and p62 were in separate disulphide-crosslinked complexes.

The 2D gel pattern obtained at a still earlier time, that is, immediately after the 2-min pulse (Fig. 1a), showed that disulphide-bonded complexes began to form before the viral glycoproteins were fully translated. In interpreting this gel, it is necessary to note that, after the short pulse, the majority of label was present in growing nascent chains of different sizes. Instead of the discrete spots given by full-length proteins (see Fig. 1b, c), nascent chains gave rise to 'spurs' running along, below or above the gel diagonal, depending on their oxidation status. The presence of spurs below the diagonal indicated that intrachain disulphide bonds were forming cotranslationally in the viral glycoproteins. Arrow A (Fig. 1a) indicates the position of the shortest nascent chains that show a deflection off the diagonal due to formation of intramolecular disulphide bonds. A fraction of nascent chains of the same length also appear to be in intermolecular complexes running above the diagonal (arrowheads), indicating that part of the labelled

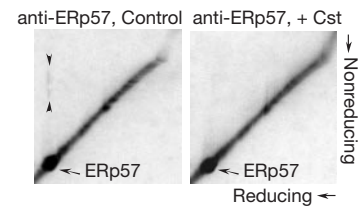


Figure 3 Erp57 is engaged in disulphide-bonded complexes with viral glycoproteins. Postnuclear supernatants of SFV-infected control cells and of cells infected in the presence of 1 mM Cst (+Cst) were precipitated with antibody to Erp57. In control cells only, a radioactive trace radiated vertically above the labelled ERp57, indicating involvement of a small fraction of the cellular oxidoreductase in intermolecular disulphide-bonded complexes. Cellular proteins were labelled in this experiment.

growing nascent glycopolypeptides formed disulphide crosslinked complexes at the same time that they started to acquire intrachain disulphide-bonds.

Thus, treatment with NEM before cell lysis, and the high level of expression in the ER of the viral glycoproteins, allowed us to trap disulphide-bonded complexes containing E1, p62 and their nascent chains. The amount of these complexes peaked after 10 min of chase (shown below) and did not exceed 0.3–1% of total labelled glycoprotein. Therefore, at any given time, only a small fraction of the newly synthesized glycoproteins was in crosslinked complexes. Quantification of the virally derived radioactivity co-immunoprecipitating with different ER chaperones showed that formation of the complexes overlapped with kinetics of glycoprotein association with calnexin (Cnx) and calreticulin (Crt) as well as with glycoprotein binding and release from the oxidoreductases PDI and Erp57 (Fig. 1e). Therefore the complexes probably constituted nascent as well as newly synthesized viral glycoproteins transiently engaged by ER oxidoreductases in mixed disulphides.

Antibodies to various ER-resident proteins (GRP94, Cnx, Crt, colligin-Hsp47, and thiol oxidases PDI, Erp57 and Erp72) were tested for their ability to precipitate from whole-cell extracts the intermolecular disulphide-bonded complexes shown in Fig. 1. The results indicated that only PDI, Erp57, Cnx and Crt were associated with, or were covalently bound components of, the complexes containing labelled viral glycoproteins (Fig. 2).

Antibodies to PDI mainly precipitated disulphide-bonded complexes visible in nonreducing gels as a ladder of labelled bands with relative molecular masses (M_r) in the 90–110K range (Fig. 2a, first lane, α PDI, -DTT). In reducing gels, the complexes dissociated, revealing that the major labelled component bonded to PDI was E1 (Fig. 2a, third lane, α PDI, +DTT). The complexes precipitating with anti-Erp57 had slower electrophoretic mobility and migrated in the 110–140K range in nonreducing gels (Fig. 2a, Erp57, -DTT). They contained both p62 and E1 as labelled components (Fig. 2a, α Erp57, +DTT). Thus, PDI and Erp57 associated with viral glycoproteins but with different specificities. p97 is a nontranslocated, misfolded, dead-end precursor of p62 and E1 that sticks nonspecifically to antibody and protein A beads; this protein was found in all immunoprecipitates as a nonspecific contaminant¹⁶.

In addition to small amounts of the disulphide-bonded complexes with electrophoretic mobility corresponding to that of Erp57 complexes (110–140K, Fig. 2a, anti-Cnx), anti-Cnx was found to precipitate non-crosslinked forms of E1 and p62. Anti-Crt also seemed to precipitate the same complexes (Fig. 2a, anti-Crt) as well as non-crosslinked p62 and a small amount of non-crosslinked E1. Precipitation with anti-Erp57, but not with anti-PDI, depleted the lysate of the disulphide-bonded complexes precipitable with anti-Cnx and anti-Crt (Fig. 2b). Colligin-Hsp47 (Fig. 2a, anti-Cng) and the oxidoreductase Erp72 (Fig. 2a, anti-Erp72) are examples of ER folding factors that did not coprecipitate with SFV glycoproteins during folding.

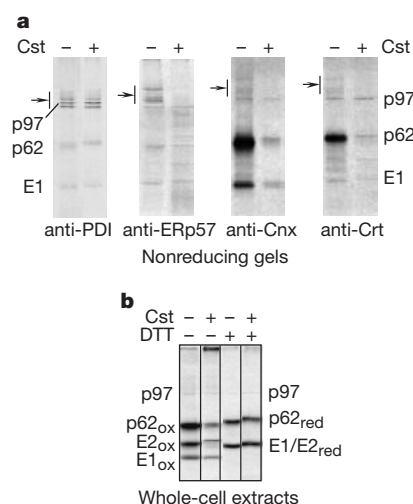


Figure 4 Inhibition of glucose trimming interferes with formation of ERp57-containing mixed disulphides and with oxidative folding of viral glycoproteins. **a**, Treatment with Cst (+Cst) efficiently inhibited the formation of ERp57-containing mixed disulphides (anti-ERp57), as well as association of viral glycoproteins with Cnx (anti-Cnx) and Crt (anti-Crt). Cst did not substantially affect formation of PDI-containing mixed disulphides (anti-PDI). Disulphide-bonded complexes are indicated with arrows. **b**, Folding status of SFV glycoproteins in control cells (lane 1, nonreducing; lane 3, reducing) and in cells treated with Cst during starvation, pulse and chase (lanes 2 and 4), in which folding efficiency substantially dropped as shown by formation of aggregates remaining at the top of the gel (second lane). Glycoproteins expressed in Cst-treated cells had slower electrophoretic mobility because their N-linked glycans remained tri-glucosylated; E1 and E2 have the same migration in the reducing gel.

Together, our results indicated that ERp57 was present in ternary complexes that contained either E1 or p62 in association with Cnx and/or Crt. That it would form complexes with these lectin-like chaperones is expected, based on studies using chemical cross-linking in microsomes and permeabilized cells^{17,18}. The complexes found between PDI and E1 did not coprecipitate with Cnx and Crt, indicating that binding of PDI to its substrate was independent of Cnx and Crt.

To label only the cellular component of the complexes, cells were incubated overnight with 0.1 mCi per dish ³⁵S-methionine and -cysteine, and infected with SFV after wash-out of radioactivity. Lysates were precipitated with antibodies to PDI, ERp57, Cnx and Crt. When analysed by 2D SDS-PAGE, ERp57-derived radioactivity was found above the diagonal. This confirmed that a small fraction of the labelled ERp57 formed intermolecular disulphide-bonded complexes (Fig. 3, anti-ERp57, control) that had a mobility in the first dimension corresponding to that of complexes containing labelled E1 and p62 (Fig. 1b, d). As E1 and p62 were the only newly synthesized proteins in the ER, the labelled ERp57 must have been disulphide-crosslinked to them. A long exposure time was needed to make the complexes visible, hence the high background radioactivity along the diagonal. The metabolic labelling of PDI was not sufficient to establish whether it was covalently associated with E1; however, it probably was, given that the covalently linked complexes had an apparent *M_r* of 90–110K and were precipitable with antibodies to both E1 (50K) and PDI (59K).

The glucosidase inhibitor castanospermine (Cst) blocked binding of E1 and p62 to Cnx and Crt (Fig. 4a, anti-Cnx and anti-Crt, +Cst). Significantly, the formation of mixed disulphides between them and ERp57 was prevented as well (Fig. 3, anti-ERp57, +Cst and Fig. 4a, anti-ERp57, +Cst). In contrast, the formation of covalent complexes containing E1 and PDI was not affected (Fig. 4a, anti-PDI, +Cst), confirming that, unlike ERp57, PDI's association with its substrate was independent of Cnx and Crt. Notably, folding of SFV glyco-

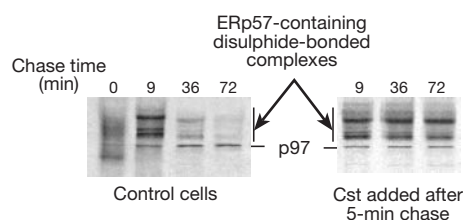


Figure 5 Glucosidase inhibitor blocks the transient nature of disulphide-bonded complexes. Postnuclear supernatants from control cells precipitated with antibody to ERp57 at different chase times show the transient nature of the ERp57-glycoproteins mixed disulphides (left panel). By blocking substrate release from Cnx and Crt, the complexes persisted beyond 72 min (right panel).

proteins was strongly perturbed under these conditions (Fig. 4b); about 80% of E1 and p62 ended up in disulphide-crosslinked aggregates that did not enter nonreducing gels (Fig. 4b, second lane, +Cst, -DTT). This implied that binding to Cnx and Crt in association with ERp57 is essential for efficient folding and proper formation of intramolecular disulphide bonds. The remaining 20% exited from the ER compartment normally, as indicated by processing of p62 into E2 (Fig. 4b, second lane, +Cst, -DTT). Thus, when the association of E1 and p62 with ERp57, Cnx and Crt was inhibited, other ER-resident chaperones or folding enzymes were able to assist the oxidative folding process, albeit less efficiently.

Delayed addition of Cst blocked the release of the glycoproteins already associated with Cnx and Crt¹⁹. This prevented proper folding and transport from the ER. The covalent complexes between ERp57 and the glycoproteins persisted beyond 72 min (Fig. 5). ERp57 was apparently trapped in complexes with the incompletely folded proteins, and perhaps caught in a futile oxidation–reduction cycle.

In summary, we have established that nascent and newly synthesized E1 and p62 formed transient complexes with ERp57 and PDI, but not with ERp72 (whether other oxidoreductases in the ER interacted similarly with newly synthesized proteins was not investigated). Association was initiated on the growing nascent chains and continued throughout the post-translational folding period. The heterogeneity and the changes in the pattern of complexes observed during chase were consistent with a reshuffling of disulphide bonds during the glycoprotein folding process. Our results show that ERp57 and PDI react directly with substrate proteins, and do not merely serve to regulate the redox state of other thiol oxidoreductases. Moreover, they underline the essential partnership between ERp57 and Cnx and Crt in oxidative folding of glycoproteins in the ER of mammalian cells. □

Received 8 July; accepted 7 September 1999.

- Gething, M.-J. & Sambrook, J. Protein folding in the cell. *Nature* **355**, 33–45 (1992).
- Helenius, A., Marquardt, T. & Braakman, I. The endoplasmic reticulum as a protein folding compartment. *Trends Cell Biol.* **2**, 227–231 (1992).
- Freedman, R. B., Hirst, T. R. & Tuite, M. F. Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem. Sci.* **19**, 331–336 (1994).
- Gilbert, H. F. Protein disulfide isomerase. *Methods Enzymol.* **290**, 27–50 (1998).
- Goldberger, R. F., Epstein, C. F. & Anfinsen, C. B. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system of rat liver. *J. Biol. Chem.* **238**, 628–635 (1963).
- Weissman, J. S. & Kim, P. S. Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature* **365**, 185–188 (1993).
- Huppa, J. B. & Ploegh, H. L. The eS-Sence of -SH in the ER. *Cell* **92**, 145–148 (1998).
- Ruddon, R. W. & Bedows, E. Assisted protein folding. *J. Biol. Chem.* **272**, 3125–3128 (1997).
- Braakman, I., Helenius, J. & Helenius, A. Role of ATP and disulphide bonds during protein folding in the endoplasmic reticulum. *Nature* **356**, 260–262 (1992).
- Krijnse-Locker, J. & Griffiths, G. An unconventional role for cytoplasmic disulfide bonds in Vaccinia virus proteins. *J. Cell Biol.* **144**, 267–279 (1999).
- Weissman, J. S. & Kim, P. S. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* **253**, 1386–1393 (1991).
- Chen, W., Helenius, J., Braakman, I. & Helenius, A. Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc. Natl Acad. Sci. USA* **92**, 6229–6233 (1995).
- Sommer, A. & Traut, R. R. Diagonal polyacrylamide-dodecyl sulfate gel electrophoresis for the identification of ribosomal proteins crosslinked with methyl-4-mercaptopbutyrimidate. *Proc. Natl Acad. Sci. USA* **71**, 3946–3950 (1974).

14. Strauss, J. H. & Strauss, E. G. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491–562 (1994).
15. Garoff, H., Simons, K. & Dobberstein, B. Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum in vitro. *J. Mol. Biol.* **124**, 587–600 (1978).
16. Marquardt, T. & Helenius, A. Misfolding and aggregation of newly synthesized proteins in the endoplasmic reticulum. *J. Cell Biol.* **117**, 505–513 (1992).
17. Oliver, J. D., van der Wal, F. J., Bulleid, N. J. & High, S. Interaction of the thiol-dependent reductase Erp57 with nascent glycoproteins. *Science* **275**, 86–88 (1997).
18. Oliver, J. D., Roderick, H. L., Llewellyn, D. H. & High, S. Erp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell.* **10**, 2573–2582 (1999).
19. Van Leeuwen, J. E. M. & Kearse, K. P. Deglycosylation of N-linked glycans is an important step in the dissociation of calreticulin-class I-TAP complexes. *Proc. Natl Acad. Sci. USA* **93**, 13997–14001 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Wileman, P. Van Nguyen and H. D. Söling for providing antibodies to Erp57 and PD1; L. Elgaard, C. Galli and A. V. Nicola for comments on the manuscript; and W. Chen and M. Corti for technical assistance. This work was supported by the ETHZ and by a Roche Foundation grant to M.M. and A.H.

Correspondence and requests for materials should be addressed to A.H. (e-mail: ari.helenius@bc.biol.ethz.ch).

The CoNR motif controls the recruitment of corepressors by nuclear hormone receptors

Xiao Hu & Mitchell A. Lazar

Division of Endocrinology, Diabetes, and Metabolism, Departments of Medicine and Genetics, and The Penn Diabetes Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

N-CoR¹ and SMRT² are transcriptional corepressors that associate with nuclear hormone receptors (NRs) in the absence of ligand. This interaction is the molecular target of differentiation therapy for acute promyelocytic leukaemia, wherein retinoic acid dissociates corepressor from leukaemogenic receptor fusion proteins^{3,4}. Binding of ligand to NRs induces a conformation that attracts coactivator proteins containing an Leu-x-x-Leu-Leu motif (the 'NR box')^{5,6}. Here we show that N-CoR and SMRT contain sequences that are similar to the NR box and are repeated in each of two NR interaction domains^{7–10}. We show that this CoNR ('corner') box is required for NR interaction, and that CoNR box peptides specifically block corepressor interaction *in vitro* and repression *in vivo*. Sequences flanking the CoNR box determine NR specificity. Thus, the key feature of hormone action, differential recognition of unliganded and liganded NRs by coactivators and corepressors, is due to very subtle differences between CoNR and NR boxes. The molecular mechanisms of repression and activation by NRs are thus linked in an unexpected manner.

Although corepressor binding requires the 'CoR box' in helix 1 (H1) of the unliganded NR¹, critical CoR box amino acids are not at the surface of the thyroid hormone receptor (TR)¹¹ and are also required for retinoid X receptor (RXR) interaction^{12,13}, suggesting that the CoR box is a structural determinant rather than a direct interaction surface. Amphipathic H12 is not required for corepressor interaction² and hinders corepressor interaction^{14,15}, but is repositioned by the ligand such that it forms part of the coactivator interaction surface along with helices H3, H4 and H5 (refs 16–19). The remainder of the NR structure is only subtly different in the absence of ligand²⁰. We therefore evaluated corepressor binding to a series of mutants in the coactivator-binding pocket of TR α (Fig. 1a) that decreased ligand-dependent activation (Fig. 1b) as predicted by

studies of TR β (ref. 16). Mutations that decreased activation also impaired or abolished repression by the unliganded receptor (Fig. 1c), whereas a mutation that did not affect activation did not alter repression or corepressor interaction (Fig. 1c, d). TR mutants with reduced repression were likewise impaired in their interactions with N-CoR and SMRT (Fig. 1d), while retaining the ability to interact with RXR (data not shown). Analogous mutants of RXR, with H12 deleted (RXR Δ AF2) to enhance corepressor interaction^{14,15}, were similarly crippled in repression (Fig. 1e) and corepressor interaction (Fig. 1f), but retained the ability to interact with TR (data not shown). Some mutants that abolished corepressor interaction, such as M1 and M4, greatly but not completely abrogated repression, suggesting a minor contribution from a different class of corepressor^{21,22}. Differences between N-CoR and SMRT and between TR and RXR were suggested by the properties of mutants M2 and M3, respectively; however, the results clearly indicated unexpected overlap between NR determinants of corepressor and coactivator interaction.

The NR coactivator interaction surface forms a hydrophobic cleft that recognizes the NR box^{16,17}. Examination of the two interaction domains (IDs) in N-CoR and SMRT revealed sequences related to the NR box that are identical in human and mouse homologues (Fig. 2a). N-CoR ID1 interacted strongly with TR but minimally with RXR, whereas ID2 interacted with both but preferred RXR (Fig. 2b), as previously described¹⁰. Mutations in the ICQII sequence of N-CoR ID1 and the LEDII sequence of ID2 were designed to test the necessity of the L/I-x-x-I/V-I sequence (Fig. 2c). Notably, mutations of both sequences abolished interaction of N-CoR with both TR and RXR Δ AF2 (Fig. 2d). The L/I-x-x-I/V-I motifs are thus required for corepressor interaction with

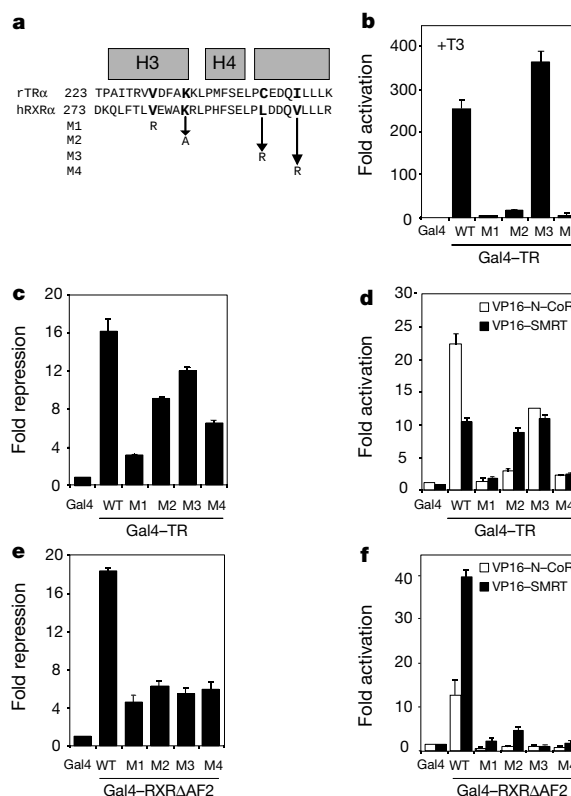


Figure 1 Corepressor interaction surface is related to the coactivator interaction surface. **a**, Sequence of helices H3, H4 and H5 in rat TR and human RXR. Bold highlights surface residues that have been mutated in M1–M4 as indicated. **b–d**, Correlation between activation (**b**), repression (**c**) and corepressor interaction (**d**; in mammalian two-hybrid assay) of Gal4–TR α . **e, f**, Mutations in the analogous amino acids of RXR Δ AF2 also decreased repression and corepressor interaction. WT, wild type.

unliganded NR. We call this motif the CoNRN ('Corepressor/nuclear receptor') box, in analogy to the NR box that is required for coactivator binding to liganded NR. Mutations in the CoNRN box sequences abolished the interactions of individual IDs (Fig. 2e, f). Mutation of L2277 to A in ID2 (m1) abolished TR interaction but preserved interaction with RXR Δ AF2. In contrast, mutation of I1 to AA (m2) eliminated interactions between ID2 and both TR and RXR Δ AF2. When examined in the context of N-CoR containing both IDs, the m2 mutation in ID1 behaved like ID2 alone (reduced interaction with TR and RXR Δ AF2), whereas the m2 mutation in ID2 behaved like ID1 alone (interaction with TR, negligible interaction with RXR Δ AF2; Fig. 2d).

Having shown that the CoNRN box was required for interaction with unliganded NR, we then tested its sufficiency. A 40-amino-acid polypeptide derived from ID1 (CoNRN1) and a 20-amino-acid ID2 polypeptide (CoNRN2) each interacted with TR *in vivo* in the absence but not the presence of ligand, recapitulating the physiology of the corepressor interaction (Fig. 3a). Consistent with the larger ID peptides, RXR Δ AF2 interacted with CoNRN2 but not CoNRN1 (Fig. 3a). Both the constitutive androstane receptor and full-length RXR, neither of which represses transcription or binds corepressors well^{14,23}, failed to interact with CoNRN1 or CoNRN2 (data not shown). The interactions involving TR were confirmed *in vitro*, as was the specificity of CoNRN1 for TR (Fig. 3b). A synthetic 20-amino-acid CoNRN2 peptide, but not a mutant or NR-box peptide

derived from the coactivator GRIP¹⁸ (Fig. 3c), prevented interaction between N-CoR and unliganded TR (Fig. 3d). However, whereas the NR box peptide inhibited GRIP1 interaction with liganded TR, neither the wild-type nor mutant CoNRN2 peptides had an effect (Fig. 3e). The CoNRN2 peptide inhibited N-CoR interaction with TR in a dose-dependent manner (Fig. 3f). The effective concentration for half-maximum response (EC₅₀) for CoNRN2 inhibition of N-CoR interaction with TR was $\sim 1 \mu\text{M}$, similar to that for NR box peptide interactions with liganded TR¹⁸. The CoNRN2 peptide, but not the mutant, also blocked interactions between N-CoR and other NRs including RXR Δ AF2, retinoic acid receptor (RAR) and the leukemogenic fusion protein PML-RAR α (Fig. 3g). Nuclear expression of the CoNRN2 peptide, but not the mutant, decreased the repression function of TR in intact cells (Fig. 3h).

We examined the molecular basis of the NR specificity of CoNRN1 and CoNRN2 by creating a series of chimaeric peptides (Fig. 4a). For TR, the amino terminus and the first amino acid in the ICQII sequence of CoNRN1 were required for interaction (chimaeras A and D, Fig. 4b). Interaction with TR was increased by combining the N terminus of CoNRN1 with the carboxy terminus of CoNRN2 (chimaera C), and the interaction was even greater when the 'x-x' was derived from CoNRN2 (chimaera B, Fig. 4b). Although the C terminus of CoNRN2 (IRKAL) resembles the $\phi\chi\phi\phi$ (where ϕ is L or I) motif, it could not substitute for the ICQII of CoNRN1 (data not shown). For RXR, the C-terminal flank

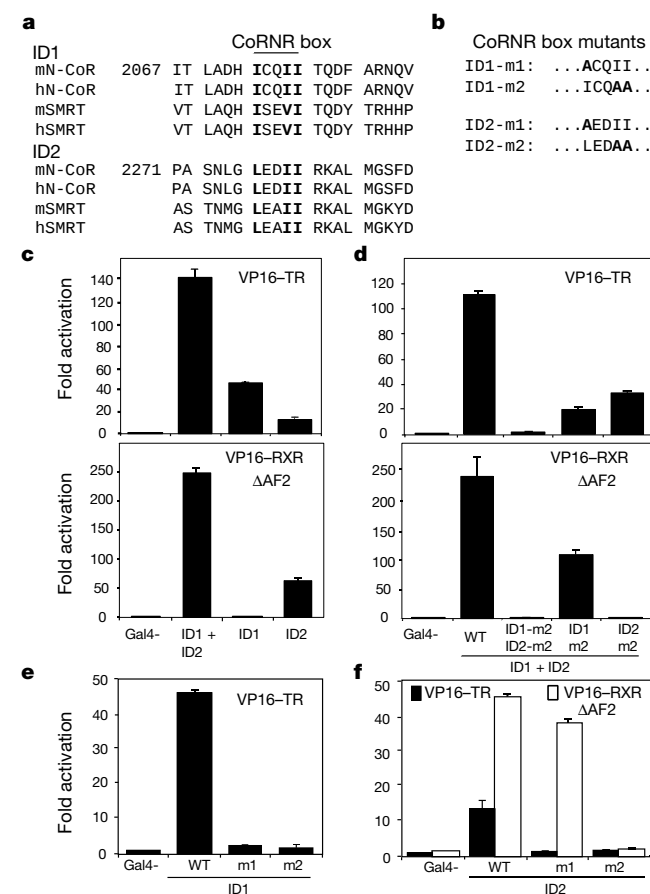


Figure 2 CoNRN boxes are necessary for interaction with NRs. **a**, Sequence alignment of the mouse and human N-CoR and SMRT IDs that contain an LxxLL-like motif. Conserved hydrophobic residues are shown in bold. **b**, Mutations in the L/I-x-x-I/V-I motif. **c**, Specificity of N-CoR IDs. ID1 + ID2: amino acids 1944–2453 of N-CoR; ID1: 1944–2239; ID2: 2239–2453. Mammalian two-hybrid assay was used to detect interactions with Gal4-TR and Gal4-RXR Δ AF2. **d**, Mutations in the L/I-x-x-I/V-I motif abrogate interactions with NRs. All mutations are in the context of ID1 + ID2. **e**, **f**, Mutation of the CoNRN box of each individual ID abolishes NR interactions.

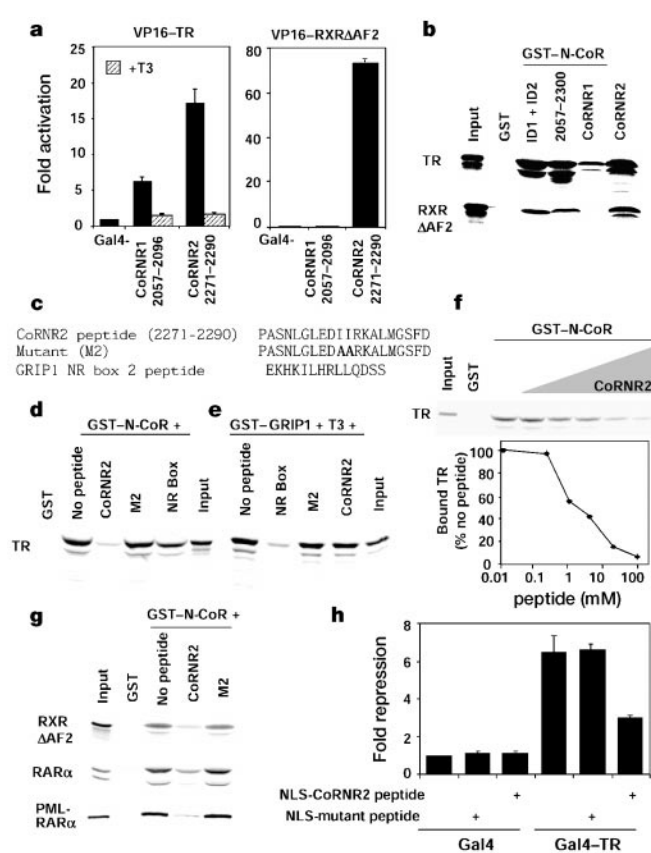


Figure 3 CoNRN box peptides are sufficient for interaction with NRs. **a**, Identification of CoNRN box peptides. **b**, *In vitro* interactions. CoNRN box peptides fused to GST were tested for interaction with TR and RXR Δ AF2. **c**, Sequences of the peptides used in **d**–**g**. Mutated residues are shown in bold. CoNRN2 and M2 peptides contain a C-terminal tyrosine. **d**, CoNRN2 but not NR box 2 peptide blocks interaction between TR and N-CoR. **e**, NR box 2 but not CoNRN2 peptide blocks interaction between TR and GRIP1. **f**, Dose response for CoNRN2 peptide. **g**, CoNRN2 peptide blocks corepressor interaction with other NRs. **h**, CoNRN2 peptide fused to the SV40 T antigen nuclear-localization signal (NLS) decreases repression by TR.

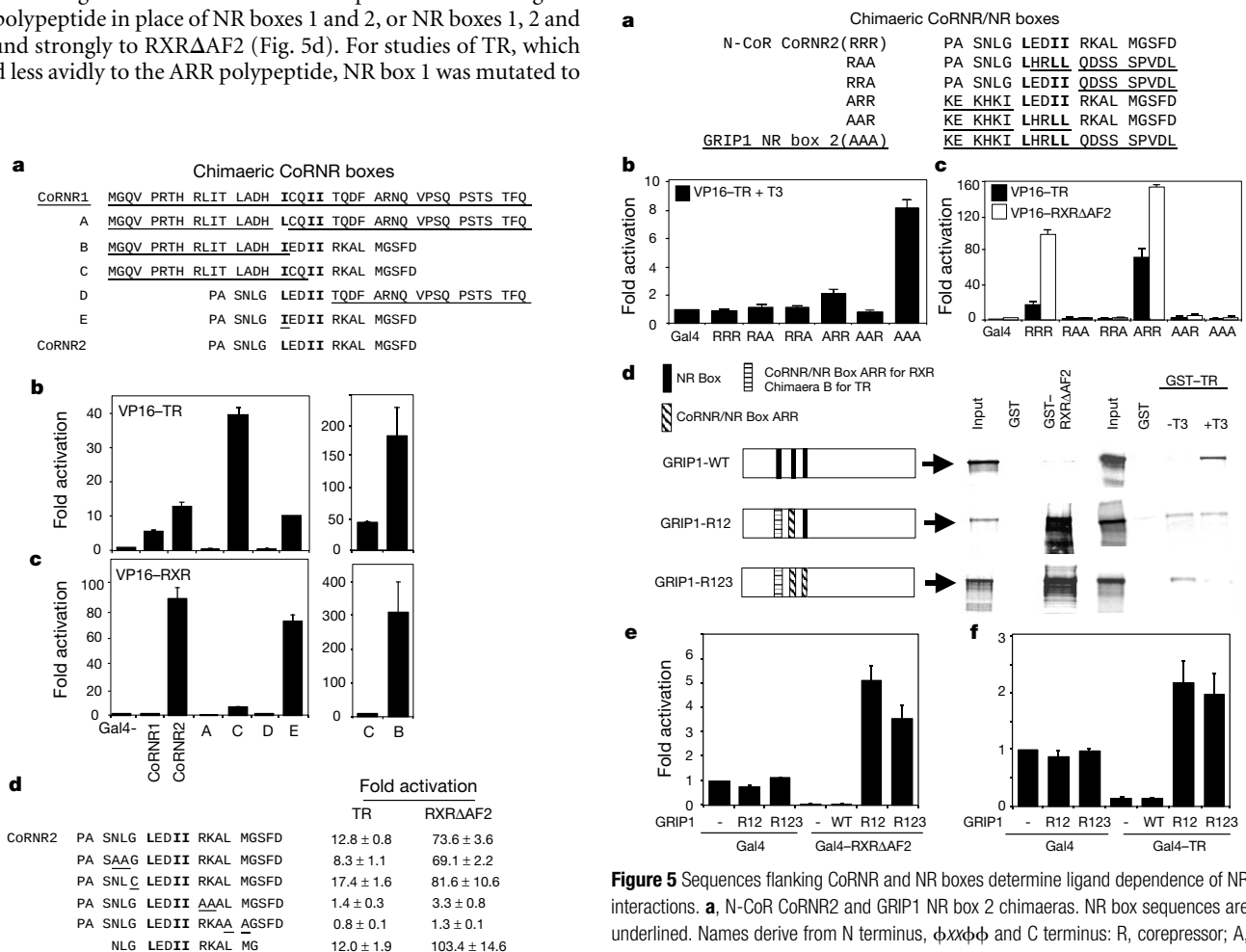
of CoNRN2 was required for interaction (Fig. 4c). These results were confirmed and extended by scanning mutations of CoNRN2 (Fig. 4d). The N-terminal flank was mutated without effect on interactions with TR and RXR, whereas the RKALM sequence immediately C-terminal to the CoNRN box was required for interaction with both TR and RXR (Fig. 4d). Moreover, the central 14 amino acids of the CoNRN2 peptide were sufficient for interaction with both TR and RXR (Fig. 4d).

We then addressed the molecular basis of the opposite ligand dependencies of NR box and CoNRN box interactions by making chimaeric polypeptides related to CoNRN2 and GRIP1 NR box 2 (Fig. 5a). The N-terminal flank, $\phi\chi\chi\phi$ and C-terminal flank were all required for ligand-dependent interaction of the NR box (Fig. 5b). This was not surprising, because structural studies have shown that amino acids from each of these regions of the NR box peptide make specific contacts with liganded TR¹⁸. The C-terminal flank of the CoNRN2 peptide was required for corepressor interaction, consistent with results shown in Fig. 4. Notably, one chimaera (ARR) containing the N-terminal flank of GRIP1 and the $\phi\chi\chi\phi$ and C terminus of CoNRN2 behaved similarly to CoNRN2 (Fig. 5c). Thus, the N terminus of the CoNRN2 peptide is not required for ligand-independent interaction.

We considered whether replacement of the GRIP1 NR boxes with CoNRN boxes would change the ligand dependence of full-length GRIP1 binding to NRs. Chimaeric GRIP1 proteins containing the ARR polypeptide in place of NR boxes 1 and 2, or NR boxes 1, 2 and 3, bound strongly to RXR Δ AF2 (Fig. 5d). For studies of TR, which bound less avidly to the ARR polypeptide, NR box 1 was mutated to

the CoNRN1/CoNRN2 chimaera B that interacted strongly with TR (Fig. 4b). As expected, binding of wild-type GRIP1 to TR required the addition of triiodothyronine (T3). Replacement of NR boxes 1 and 2 with CoNRN boxes enhanced interaction in the absence of T3, although some binding was also observed in the presence of T3, perhaps due to NR box 3. When all three NR boxes were replaced with CoNRN boxes, binding of GRIP1 was ligand independent, and clearly reduced by T3 (Fig. 5d). Thus, the CoNRN boxes reversed the normal ligand dependence of GRIP1, and converted the NR interactions of this coactivator to those characteristic of the corepressors.

Coactivators are modular, with separate transcription-activation domains that are distinct from the NR boxes^{24,25}. Thus, it was possible that the altered ligand dependence of the CoNRN box substitutions might convert GRIP1 from a ligand-dependent coactivator to a ligand-independent coactivator. Indeed, the chimaeric GRIP1 not only reversed repression, but also activated transcription by RXR (Fig. 5e), as well as TR (Fig. 5f), well beyond the 'basal' transcription observed with Gal4 alone. Thus, CoNRN box substitutions transformed GRIP1 into a ligand-independent coactivator, thereby reversing its normal physiological function. The capacity of a relatively subtle amino-acid substitution to transform GRIP1 into a coactivator for unliganded NRs confirms the importance of the NR recognition code in a cellular context.



We have shown that coactivators and corepressors use similar mechanisms of interaction with distinct NR conformations. Our results begin to shed light on a code underlying NR interactions with functionally distinct coregulators. Sequences flanking both sides of the coactivator LxxLL core impart receptor specificity^{18,26,27} and dictate that coactivators exclusively recognize the ligand-bound NR conformation. Corepressor CoRNR1 uses a unique sequence on the N terminal side of its I-x-x-I/V-I core for binding aporeceptors, whereas CoRNR2 interaction with the unliganded NR conformation requires specific C-terminal sequences flanking an LxxII core. The precise sequences of the ϕ x ϕ cores and flanking sequences are critical determinants of receptor preference. Differences between N-CoR and SMRT may explain why TR preferentially interacts with N-CoR, whereas RXR interacts better with SMRT (Fig. 1d, f). Furthermore, the presence of two CoRNR boxes per corepressor is probably related to the stoichiometry of NR binding²⁸, and the receptor specificity of CoRNR1 and CoRNR2 may confer specificity for NR homo- and heterodimers^{9,10,13}. The ability of short CoRNR polypeptides to selectively block corepressor or coactivator interaction with NR specificity indicates that peptidomimetic compounds may be developed with NR-selective abilities to block repression or activation. These could be of therapeutic benefit in syndromes of thyroid hormone resistance²⁹ and acute promyelocytic leukaemia^{3,4}. □

Methods

Constructs

Deletion constructs and chimera constructs were made by PCR or ligation of double-stranded oligonucleotides. Point mutations were made using the Quick-change site-directed mutagenesis kit (Stratagene) or double-stranded oligonucleotides. We completely sequenced all relevant regions.

Glutathione S-transferase (GST)-pull-down assay

GST-pull-down experiments were done as described²⁸. Peptides used in the GST-pull-down experiments were synthesized by Cybersyn or the Protein Chemistry laboratory of the University of Pennsylvania Medical Center (supported by the core grant of the Penn Diabetes Center). Competing peptides were dissolved in DMSO and used at 100 μ M unless otherwise indicated. Quantification was performed by phosphorimaging. T3 concentration was 10 μ M.

Transient transfection assays

Transient transfections assays of repression, activation and mammalian two-hybrid interactions were performed as described¹³ using lipofectamine (GIBCO BRL). All transfections were done in 293T cells. All Gal4 experiments used a (Gal4 $\times$ 5)-SV40-Luc reporter and a control β -galactosidase (β -gal) construct. T3 concentration was 1 μ M.

Received 18 June; accepted 23 September 1999.

- Horlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397–404 (1995).
- Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454–457 (1995).
- Lin, R. J. *et al.* Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* **391**, 811–814 (1998).
- Grignani, F. *et al.* Fusion proteins of the retinoic acid receptor- α recruit histone deacetylase in promyelocytic leukaemia. *Nature* **391**, 815–818 (1998).
- LeDourain, B. *et al.* A possible involvement of TIF1 α and TIF1 β in the epigenetic control of transcription by nuclear receptors. *EMBO J.* **15**, 6701–6715 (1996).
- Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733–736 (1997).
- Seol, W., Mahon, M. J., Lee, Y.-K. & Moore, D. D. Two receptor interacting domains in the nuclear hormone receptor corepressor RIP13/N-CoR. *Mol. Endocrinol.* **10**, 1646–1655 (1996).
- Zamir, I. *et al.* A nuclear hormone receptor corepressor mediates transcriptional silencing by receptors with different repression domains. *Mol. Cell. Biol.* **16**, 5458–5465 (1996).
- Cohen, R. N., Wondisford, F. E. & Hollenberg, A. N. Two separate NCoR (nuclear receptor corepressor) interaction domains mediate corepressor action on thyroid hormone response elements. *Mol. Endocrinol.* **12**, 1567–1581 (1998).
- Wong, C. W. & Privalsky, M. L. Transcriptional silencing is defined by isoform and heterodimer-specific interactions between nuclear hormone receptors and corepressors. *Mol. Cell. Biol.* **18**, 5724–5733 (1998).
- Wagner, R. L. *et al.* A structural role for hormone in the thyroid hormone receptor. *Nature* **378**, 690–697 (1995).
- Collingwood, T. N. *et al.* Thyroid hormone-mediated enhancement of heterodimer formation between thyroid hormone receptor β and retinoid X receptor. *J. Biol. Chem.* **272**, 13060–13065 (1997).
- Zhang, J., Zamir, I. & Lazar, M. A. Differential recognition of liganded and unliganded thyroid hormone receptor by retinoid X receptor regulates transcriptional repression. *Mol. Cell. Biol.* **17**, 6887–6897 (1997).

- Zhang, J., Hu, X. & Lazar, M. A. A novel role for helix 12 of RXR in regulating repression. *Mol. Cell. Biol.* **19**, 6448–6457 (1999).
- Schulman, I. G., Juguilon, H. & Evans, R. M. Activation and repression by nuclear hormone receptors: hormone modulates an equilibrium between active and repressive states. *Mol. Cell. Biol.* **16**, 3807–3813 (1996).
- Feng, W. *et al.* Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors. *Science* **280**, 1747–1749 (1998).
- Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- γ . *Nature* **395**, 137–143 (1998).
- Darimont, B. D. *et al.* Structure and specificity of nuclear receptor–coactivator interactions. *Genes Dev.* **12**, 3343–3356 (1998).
- Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
- Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384–391 (1998).
- Dressel, U. *et al.* Alien, a highly conserved protein with characteristics of a corepressor for members of the nuclear hormone receptor superfamily. *Mol. Cell. Biol.* **19**, 3383–3394 (1999).
- Zamir, I. *et al.* Cloning and characterization of a corepressor and potential component of the nuclear hormone receptor repression complex. *Proc. Natl Acad. Sci. USA* **94**, 14400–14495 (1997).
- Forman, B. M. *et al.* Androstane metabolites bind to and deactivate the nuclear receptor CAR β . *Nature* **395**, 612–615 (1998).
- Onate, S. A. *et al.* The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors. *J. Biol. Chem.* **273**, 12101–12108 (1998).
- Voegel, J. J., Heine, M. J. S., Zechel, C., Chambon, P. & Gronemeyer, H. TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.* **15**, 3667–3675 (1996).
- McInerney, E. M. *et al.* Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev.* **12**, 3357–3368 (1998).
- Mak, H. Y., Hoare, S., Henttu, P. M. & Parker, M. G. Molecular determinants of the estrogen receptor–coactivator interface. *Mol. Cell. Biol.* **19**, 3895–3903 (1999).
- Zamir, I., Zhang, J. & Lazar, M. A. Stoichiometric and steric principles governing repression by nuclear hormone receptors. *Genes Dev.* **11**, 835–846 (1997).
- Yoh, S. M., Chatterjee, V. K. K. & Privalsky, M. L. Thyroid hormone resistance syndrome manifests as an aberrant interaction between mutant T3 receptors and transcriptional corepressors. *Mol. Endocrinol.* **11**, 470–480 (1997).

Acknowledgements

We thank J. Zhang for helpful discussions, and D. Moore for providing VP16-CAR. This work was supported by grants from the National Institute of Diabetes, Digestive and Kidney Diseases of the NIH to M.A.L.

Correspondence and requests for materials should be addressed to M.A.L. (e-mail: lazar@mail.med.upenn.edu).

Mitochondrial DNA repairs double-strand breaks in yeast chromosomes

Miria Ricchetti*, Cécile Fairhead† & Bernard Dujon†

* Unité de Physicochimie des Macromolécules Biologiques (URA1773 du CNRS), Institut Pasteur, 25–28 Rue du Docteur Roux, 75724 Paris Cedex 15, France
† Unité de Génétique Moléculaire des Levures (URA1300 du CNRS, UFR 927 Université Pierre et Marie Curie), Institut Pasteur, 25–28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

The endosymbiotic theory for the origin of eukaryotic cells¹ proposes that genetic information can be transferred from mitochondria to the nucleus of a cell, and genes that are probably of mitochondrial origin have been found in nuclear chromosomes². Occasionally, short or rearranged sequences homologous to mitochondrial DNA are seen in the chromosomes of different organisms including yeast, plants and humans³. Here we report a mechanism by which fragments of mitochondrial DNA, in single or tandem array, are transferred to yeast chromosomes under natural conditions during the repair of double-strand breaks in haploid mitotic cells. These repair insertions originate from non-contiguous regions of the mitochondrial genome. Our analysis of the *Saccharomyces cerevisiae* mitochondrial genome⁴ indicates that the yeast nuclear genome does indeed contain several short sequences of mitochondrial origin which are similar in size and

composition to those that repair double-strand breaks. These sequences are located predominantly in non-coding regions of the chromosomes, frequently in the vicinity of retrotransposon long terminal repeats, and appear as recent integration events. Thus, colonization of the yeast genome by mitochondrial DNA is an ongoing process.

In the yeast *S. cerevisiae*, 90% of the repair of double-strand breaks (DSBs) in chromosomes occurs by homologous recombination⁵. Non-homologous events, which account for the remaining repair events, typically involve the re-ligation of DNA ends by occasional base-pairing⁶⁻⁷. Rarely (about 1% of non-homologous events), repair occurs by insertion of a short retrotransposon sequence⁸⁻⁹.

We have investigated non-homologous repair events in haploid *S. cerevisiae* cells after induction of a DSB, mediated by the *I-SceI* nuclease. A *URA3* cassette, between two *I-SceI* sites in opposite orientation, was introduced into chromosome XI in place of either *YKL222c* (D. Alexandraki, unpublished data) or *YKR098c* (ref. 10). Haploid cells were then transformed with the plasmid pPEX7, which bears the *LEU2* marker gene and which encodes the rare cutter restriction endonuclease *I-SceI* under the control of a galactose-inducible promoter¹¹. Transformed cells were plated onto a medium selective for leucine and containing galactose as the sole carbon source. Expression of *I-SceI* resulted in the removal of the *URA3* gene (Fig. 1). Under these conditions, repair by homologous recombination was not possible and the great majority of cells did not form colonies. Approximately 1% of the total population formed colonies, compared with control cells (grown on the same medium but containing glucose, and thus not expressing *I-SceI*). Polymerase chain reaction (PCR) fragments of 1,248 and 894 base pairs (bp) were expected if re-ligation of the extremities occurs by non-homologous end-joining (NHEJ) at the *YKL222c* and the *YKR098c* loci, respectively (Fig. 1). Out of a total of 265 randomly isolated survivors analysed, 1 (0.4%) resulted in a shorter and 5 (1.9%) in a longer DNA fragment, and the remaining clones either gave PCR fragments of the expected length or did not amplify (see Table 1, experiment 1). Sequencing of seven randomly chosen DNA fragments of the expected length showed that re-ligation of the extremities by NHEJ had occurred, as previously described⁷. The

shorter DNA fragment had a deletion of 144 bp (28 and 116 bp from each extremity, respectively).

The five longer DNA fragments all contained an insertion which corresponded to mitochondrial DNA sequences (Fig. 2). In survivors 34pAS7, 34pAS16 and 34pAT9 (all corresponding to repair at the *YKR098c* locus), single insertions of 47, 77 and 97 bp fragments of mitochondrial DNA, respectively, had occurred. These insertions exhibited sequences identical to three distinct regions of the mitochondrial map (Fig. 3b), and all showed a concomitant loss of a single A residue at one of the two original DNA extremities. The two remaining survivors analysed, 622pBS8 (repair at the *YKL222c* locus) and 34pAS15 (repair at the *YKR098c* locus), were more complex because, in both cases, two non-contiguous mitochondrial sequences were inserted in a tandem array at each locus (Fig. 2). Survivor 622pBS8 contained two DNA fragments of 101 and 52 bp, with 100% identity to mitochondrial regions 19870–19770 and 30114–30063, respectively. Both the proximal and the distal nuclear chromosomal moieties were deleted by one nucleotide (Fig. 2). Survivor 34pAS15 contained two DNA fragments of 189 and 49 bp with 100% identity to mitochondrial regions 74600–74412 and 34946–34994, respectively. This insertion resulted in an intact chromosomal sequence proximally and a 4-nucleotide chromosomal deletion distally. A second anomaly of this survivor was the presence of one G residue downstream of the double insertion (Fig. 2). This residue could originate either from a single base addition concomitant to the repair process or from a point mutation. Sequencing or PCR artifacts were excluded as this residue was detected in three independent experiments and was absent from two independent mitochondrial DNA sequencing studies^{4,12}.

The inserted mitochondrial sequences described corresponded to coding and non-coding mitochondrial DNA, and to both orientations with respect to the major transcribed strand of the mitochondrial genome⁴ (see Fig. 3). One of the two mitochondrial sequences inserted in 34pAS15 is part of a G-C cluster. All other insertions have the typical A-T-rich pattern of yeast mitochondrial DNA. In all of the five cases in experiment 1, deletion of one residue (four

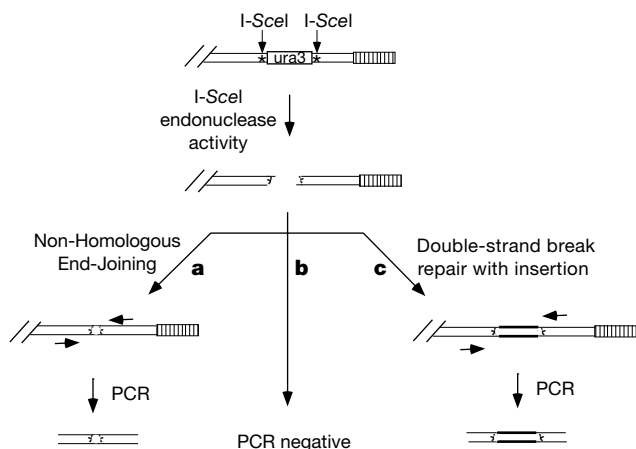


Figure 1 Flow-diagram of the experimental design. An artificial *URA3* cassette between two *I-SceI* sites is inserted in yeast chromosome XI replacing either *YKL222c* or *YKR098c*. The scheme represents part of the right arm of chromosome XI (where the deleted ORF *YKR098c* was located; a mirror symmetry has to be used for the deleted ORF *YKL222c* which is located on the left arm of the chromosome). Striped rectangles indicate the telomere. Asterisk indicates an intact *I-SceI* site. After action of *I-SceI*, the double-strand break (DSB) can be repaired by either non-homologous end-joining (a) or by insertion of a foreign sequence (c). The two mechanisms are distinguished by PCR amplification from oligonucleotides located in flanking sequences (arrows). PCR negative cases (b) are not considered in this work.

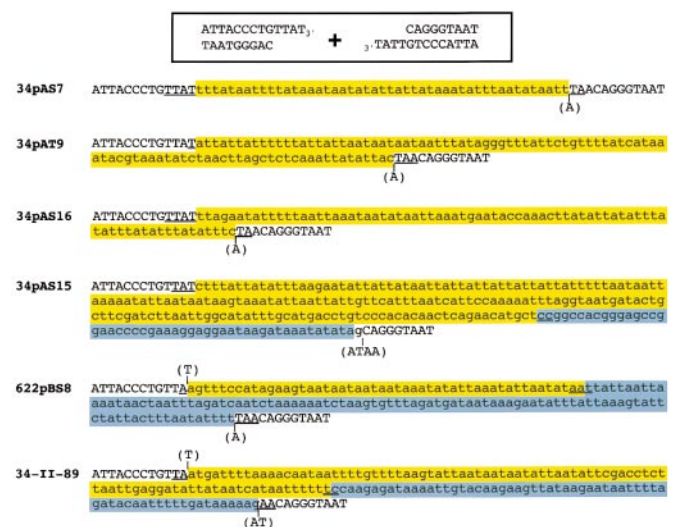


Figure 2 Sequencing of repaired double-strand breaks. The box at the top of the figure indicates expected chromosome ends resulting from *I-SceI* cleavage (the two *I-SceI* sites were placed in opposite orientation). Chromosomal sequences are indicated by capital letters (deleted nucleotides are shown in parentheses) and inserted sequences by lower-case letters. Coloured sequences are identical to mitochondrial DNA. Non-contiguous mitochondrial sequences are represented by different colours. Underlined residues indicate sequence homology at the ends of fragments either between the insertion and the chromosomal moiety or, in the case of double insertions, between the two inserted mitochondrial DNA fragments. All sequences are shown oriented as in the Watson strand of chromosome XI.

Table 1 PCR analysis of survivors after DSB induction

Experiment	Relevant nuclear genotype	Mitochondrial genotype	Total number of tested clones	Longer PCR DNA fragments
1	$\Delta ykl222c::URA3-I-SceI$	ρ^+	92	1
1	$\Delta ykr098c::URA3-I-SceI$	ρ^+	173	4
2	$\Delta ykr098c::URA3-I-SceI$	ρ^+	166	1
2	$\Delta ykr098c::URA3-I-SceI$	ρ^-	177	0
2	$\Delta ykr098c::URA3-I-SceI$	ρ^0	173	0
2	$\Delta ykr098c::URA3-I-SceI$	ρ^0	47	0
2	$\Delta ykr098c::URA3-I-SceI$	ρ^0	165	0
2	$\Delta ykr098c::URA3-I-SceI, \Delta yme1::KANMX$	ρ^+	240	0

All strains are FYBL2-5D derivatives transformed with the I-SceI expression vector pPEX7. Complete genotypes are given in the Methods; relevant aspects are indicated in column 2 or 3 for the nuclear or mitochondrial genomes, respectively.

residues in 34pAS15) at the distal end of the chromosomal DSB suggests that a specific excision activity accompanies the mitochondrial DNA insertion. Based on the complete mitochondrial sequence⁴, the two fragments in the double insertion events share a terminal homology of 2–3 nucleotides (underlined in Fig. 2). A homology of 1–4 nucleotides is also present between the inserted sequences and the expected DSB ends. The presence of such short homologies is typical of non-homologous junctions in mammalian¹³ and yeast cells⁶.

The insertion of ‘filler’ DNA, usually less than 40 bp, is common at illegitimate junctions in mammalian and plant cells, where it is accompanied by rearrangements at the junction site^{14–16}. In contrast, our analysis shows that the integration of mitochondrial sequences into the *S. cerevisiae* chromosome is not accompanied by modifications at the junction site, except for the loss or gain of 1–4 nucleotides (Fig. 2). This was also the case for the insertion of retrotransposon Ty1 DNA fragments in yeast⁹. This suggests that the mechanism of integration of exogenous sequences into yeast chromosomes differs from that of filler DNA in higher eukaryotes.

Transfer of mitochondrial sequences to the nucleus has been previously reported in yeast after transformation¹⁷ or after micro-projectile bombardment¹⁸. In the latter case, artificial ρ^- mutants, which did not contain endogenous mitochondrial DNA but carried a plasmid bearing a mitochondrial and a nuclear marker gene, increased the transfer efficiency by ~30 times compared with normal ρ^+ cells (containing a wild-type mitochondrial genome)¹⁹. Similarly, the nuclear mutation *yme1* increased transfer in ρ^+ cells¹⁹.

To determine whether ρ^- and *yme1* mutants would increase the frequency of DSB repair by mitochondrial DNA, we constructed a ρ^- derivative and *yme1* mutant of the $\Delta ykr098c$ strain and repeated DSB induction and analysis of the surviving clones (Table 1, experiment 2). Our results show that no single case of transfer was observed for the ρ^- mutant (0/177 survivors) or for *yme1* (0/240 survivors). Additional experiments, performed with three different ρ^0 mutants, whose mitochondria do not contain DNA, did not show any mitochondrial insertion (0/385 survivors). As a control, when the experiment was repeated with the original $\rho^+ YME1$ ($\Delta ykr098c$) strain, another case of transfer was found (1/166 survivors). This survivor, designated 34-II-89, contained a double insertion of 92 and 65 bp having 100% identity to mitochondrial regions 23396–23487 and 14804–14868, respectively (Figs 2 and 3). These results demonstrate that some mitochondrial DNA fragments are transferred to the nucleus of normal cells, at least transiently. In contrast to a previous report¹⁹, this transfer appears to be independent of mitochondrial genomic alterations and of the *yme1* mutation.

The mechanisms by which mitochondrial DNA molecules escape the organelle and reach the nucleus remain to be elucidated. Membrane fusion, lysis of mitochondria, and illegitimate use of the nuclear import machinery have all been proposed (see ref. 3). Given that in yeast the nuclear membrane remains intact during mitosis²⁰, the passive incorporation of exogenous nucleic acids into the nucleus during this phase is unlikely.

The frequency of the integration events observed here (~1.5% of

survivors) suggests that this is not an exceptional event. We therefore re-examined the presence of mitochondrial sequences in the nuclear genome by comparing the sequence of the sixteen yeast chromosomes with the recently sequenced mitochondrial genome⁴ (both sequences are from the same strain). Table 2 lists the 30 mitochondrial sequences (22–230 bp in length) that were found in 13 out of 16 yeast chromosomes, with sequence identity varying from 86–100%; a few of these sequences were previously described^{21–23}, but 22 are new. Among these, a 123-bp mitochondrial intergenic segment in chromosome IV, a 104-bp segment of

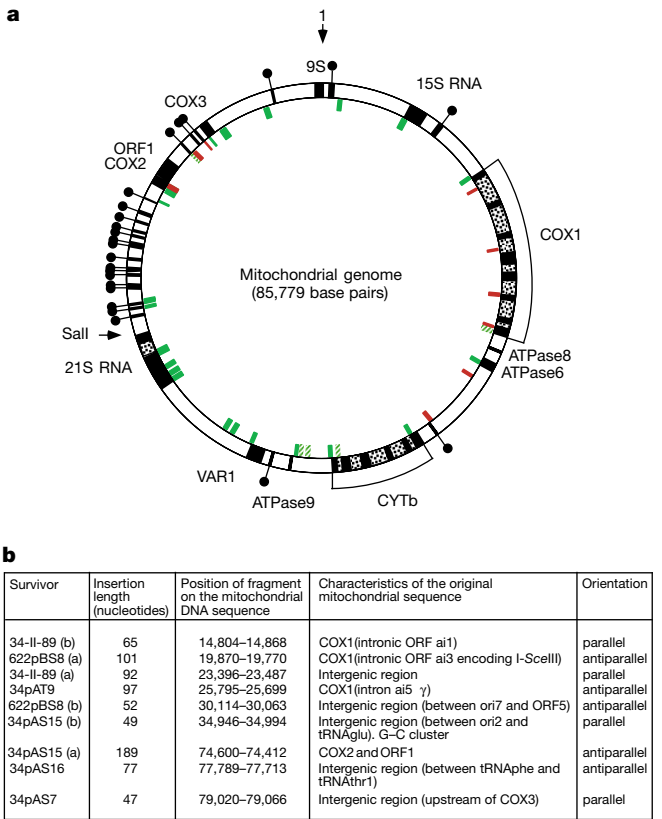


Figure 3 Description of mitochondrial DNA found in the nucleus. **a**, Mitochondrial map. The map of *Saccharomyces cerevisiae* mitochondrial DNA is deduced from its complete sequence⁴. Filled boxes indicate coding exons, dotted boxes indicate introns. Only major mitochondrial genes are labelled; transfer RNA gene locations are indicated by sticks with circles (see ref. 4 for complete annotation). Segments of mitochondrial DNA sequences found in the nuclear yeast sequences are indicated by green boxes (see Table 2 for complete list); those found more than once are striped. Segments of mitochondrial DNA transferred to the nucleus after DSB repair are indicated as red boxes. **b**, Description of the sequences inserted at the DSB. The orientations ‘parallel’ (or ‘antiparallel’) indicate that the sequence of the Watson strand of chromosome XI (Fig. 2) corresponds (or does not correspond) to the non-transcribed strand of mitochondrial DNA.

Table 2 Mitochondrial sequences present in *S. cerevisiae* chromosomes

Mitochondrial segment						Nuclear segment		
Coordinates	Gene/intergene	Orientation	Size (bp)	Percentage identity	BLAST2 score	Chromosome number	Chromosome coordinates (reference)	Chromosome region
41088-41317	* CYTB-I4	<	230	100	1.9×10^{-89}	IX	I 7245-7016 (ref. 22)	YIL177c <> YIL176c
41088-41317	* CYTB-I4	<	230	100	1.9×10^{-89}	X	J 7228-6999 (ref. 22)	YJL225c <> YJL223c
1367-1489	tRNAPro <> ORF6	<	123	99	3.3×10^{-44}	IV	D 544798-544676	YDR043c <> YDR044w
61906-62037	21SRNA <> tRNA thr2	<	132	86	2.1×10^{-16}	II	B 363938-363807 (ref. 23)	YBR060c <> YBR061c
6836-6889	15S RNA	>	54	100	2.9×10^{-16}	XII	L 682556-682609	YLR270w <> YLR271w
58961-59016	21S RNA	<	56	98	6.3×10^{-16}	XI	K 16248-16193 (ref. 23)	YKL219w <> YKL218c
42749-42800	* CYTB <> ori6	>	52	98	1.1×10^{-14}	I	A 189215-189266 (ref. 23)	YAR033w <> LTR-YARdelta7
42749-42800	* CYTB <> ori6	>	52	96	6.3×10^{-14}	VII	G 404800-404851	YGL051w <> LTR-YGLWdelta8
36933-37032	CYTB (b12 + non-coding)	<	104	92	1.3×10^{-13}	X	J 416522-416419	LTR-YJLWdelta8 <> LTR-YJLWdelta9
74289-74349	COX2	<	61	86	3.5×10^{-13}	IX	I 429527-429467 (ref. 23)	YIR038c <> YIR039c
42828-42909	* CYTB <> ori6	>	83	98	6.3×10^{-13}	VII	G 404849-404931	YGL051w <> LTR-YGLWdelta8
42828-42909	* CYTB <> ori6	>	83	98	6.3×10^{-13}	I	A 189264-189346 (ref. 23)	YAR033w <> LTR-YARdelta7
43605-43648	CYTB <> ori6	>	44	98	2.4×10^{-11}	XIII	M 183458-183501 (ref. 21)	YML046w <> tRNA Gly
49454-49530	VAR1	>	78	95	1.4×10^{-10}	XIII	M 183523-183600 (ref. 21)	YML046w <> tRNA Gly
73712-73792	tRNA met <> COX2	>	78	92	6.7×10^{-10}	IX	I 246046-246123 (ref. 23)	YIL061c <> LTR-YILCdelta2
29100-29169	ATPase6	>	70	95	1.9×10^{-09}	XII	L 412638-412707	YLR134w <> YLR135w
58341-58384	21S RNA	<	44	90	4.2×10^{-09}	IV	D 159237-159194	YDL169c <> YDL168w
26324-26359	* COX1	<	36	100	9.0×10^{-09}	VII	G 404972-404937	YGL051w <> LTR-YGLWdelta8
64412-64494	includes tRNAcys	<	83	88	1.4×10^{-08}	VII	G 2350-2268	LTR-YGLomega1 <> YGL263w
58841-58922	21S RNA	<	92	86	8.0×10^{-08}	VII	G 567910-567819	LTR-YGRCdelta16 <> LTR-YGRCdelta17
6556-6593	15S RNA	<	38	92	2.3×10^{-07}	XIII	M 57783-57746	YML106w <> YML105c
26330-26359	* COX1	<	30	100	2.8×10^{-06}	I	A 189382-189353 (ref. 23)	YAR033w <> LTR-YARdelta7
45799-45835	ori6 <> ATPase9	>	37	89	3.4×10^{-06}	VII	G 897226-897262	YGR198w <> YGR199w
79268-79326	COX3	<	59	88	2.4×10^{-05}	III	C 272219-272161 (ref. 23)	YCR090c <> YCR091w
60312-60338	21S RNA	<	27	100	5.0×10^{-05}	XV	O 462258-462232	YOR072w <> YOR073w
13770-13799	ori8 <> COX1	<	30	93	8.8×10^{-05}	XV	O 1037531-1037560	within YOR373w
64614-64643	tRNAhis	<	30	93	8.8×10^{-05}	VII	G 2219-2190	LTR-YGLomega1 <> YGL263w
52679-52702	ORF10 <> ori3	<	24	100	8.8×10^{-04}	XVI	P 939132-939155	YPR199c <> YPR200c
51903-51927	ORF10 <> ori3	<	25	96	1.9×10^{-03}	XI	K 428468-428444	YKL008c <> YKL007w
64603-64627	tRNAhis	<	25	96	1.9×10^{-03}	XVI	P 52950-52926	YPL259c <> YPL258c
84295-84319	ori5 <> tRNAf-met	<	25	96	1.9×10^{-03}	VIII	H 197263-197239	YHR045w <> YHR046c
80378-80400	COX3 <> ori5	<	23	100	2.3×10^{-03}	VII	G 450114-450092	within YGL024w
77611-77634	tRNAphe <> tRNAthr1	<	24	95	4.9×10^{-03}	XVI	P 777107-777084	LTR-YPRCdelta15 <> LTR-YPRWdelta16
13723-13744	ori8 <> COX1	>	22	100	6.0×10^{-03}	XII	L 35427-35448	YLL054c <> YLL053c

Results are classified by decreasing order of probabilities compared with random occurrence of similar sequences. Mitochondrial sequences that have undergone substantial divergence or which have a short size are not reliably detectable and were therefore excluded from this analysis. The mitochondrial and nuclear chromosomal coordinates of the homologous sequences are given, together with the positions relative to genes on the mitochondrial and chromosomal maps, where the symbol <> indicates genomic objects flanking the analysed sequence. Nomenclature of mitochondrial genetic objects, such as ori, ORF, cox, is as defined in ref. 4. Sequences previously observed are indicated with a reference number on the right of the chromosome coordinates. The size of the nuclear sequence is indicated, as is its percentage of identity with the corresponding mitochondrial sequence. The orientation of the mitochondrial segment is given, where > indicates that the non-transcribed strand of the mitochondrial genome is represented on the Watson strand of the nuclear yeast sequence, and < the Crick strand. Mitochondrial sequences that are present more than once in yeast chromosomes are indicated with an asterisk.

†, ‡, § Multiple insertions of mitochondrial sequences at the same chromosomal site are indicated with identical symbols.

|| Chromosomal regions that are part of ancestral duplications²⁸.

the *CYTB* gene (containing both coding and non-coding regions) in chromosome X, a fragment of 92 bp of the mitochondrial 21S RNA gene and a region of 83 bp which includes mitochondrial tRNAcys gene both in chromosome VII and a fragment of 70 bp of *ATPase6* in chromosome XII. The intergenic region of 83 bp, previously seen only in chromosome I (ref. 23), is also present in chromosome VII. In most cases, a given mitochondrial fragment is found only once in the nuclear genome. However, four sequences are present twice, and are located on different chromosomes (asterisks in Table 2). These have homologous or identical flanking sequences for several thousand bp^{22,24,25}, suggesting that they result from duplication of chromosomal regions that occurred after the mitochondrial DNA insertion.

In a few cases (see chromosomes I, VII, and XIII, Table 2) multiple insertions of non-contiguous mitochondrial fragments are present at the same site on the chromosome, as in the cases of the double insertions found in our DSB repair experiments. In a previous study²¹, the authors postulated that such non-contiguous sequences may have originated from ρ^- rearranged mitochondrial DNA transferred to the nucleus. However, the survivors described in our experiments were ρ^+ (data not shown) and the ρ^- mutant failed to repair DSBs with its mitochondrial DNA (experiment 2). Although we cannot exclude the possibility that rare ρ^- mitochondria, present in wild-type ρ^+ cells, could be a source of deleted and/or rearranged DNA, we think that the joining of two mitochondrial sequences observed here occurred during the repair of DSBs.

A total of 34 distinct insertion sites, including the double events, was observed. Ninety-four per cent (32/34) of the mitochondrial sequences are inserted within intergenic regions of the yeast genome and, interestingly, 44% (14/32) of these are in the immediate vicinity of retrotransposon long terminal repeats (LTRs) or transfer

RNA (tRNA) genes (Table 2). Because intergenic regions represent only 28% of the *S. cerevisiae* nuclear genome²⁶, traces of mitochondrial DNA sequences are 41 times more frequent in the non-coding areas than in the coding ones, consistent with the notion that integration of mitochondrial DNA into a gene would be deleterious. Alternatively, mitochondrial DNA sequences may preferentially insert into non-coding regions after, for example, transient DSBs occurring during meiosis. However, the insertion in Table 2 and other shorter sequences described in the Methods do not correlate with the mapped 'hotspots' of meiotic recombination in chromosome III (ref. 27). Table 2 and Fig. 3a show that the transferred sequences originate from all regions of the mitochondrial DNA map. They correspond to coding (38%) and non-coding areas (62%) in proportions close to the relative proportions of these regions in the mitochondrial genome, indicating an absence of bias for the origin of the colonizing DNA.

The complete or near-complete identity of the chromosomal sequences with mitochondrial DNA of the same strain, and the fact that when paralogous chromosome regions (resulting from an ancestral duplication event) are present²⁸ mitochondrial sequences are found in only one of them, indicate that these events correspond to recent transfers.

Thus, DSB repair offers a plausible mechanism to explain mitochondrial insertions in nuclear chromosomes and to account for the presence of mitochondrial DNA fragments and even genes in nuclear genomes. Mitochondrial transfer can take place under laboratory conditions, but it has certainly occurred naturally. From our experiments and from the analysis of the yeast genome sequences, we conclude that a constant flow of genetic information takes place from mitochondria to the nucleus. The precise mechanisms of mitochondrial insertions, which affects the evolution of the nuclear genome, need to be further elucidated. Unresolved

questions concern the mechanism of transfer, the frequent co-insertion of non-contiguous mitochondrial inserts and the presence of mitochondrial insertions in the vicinity of LTRs and tRNA genes. □

Methods

Experimental methods

The two haploid strains used are derivatives of FYBL2-5D (*MAT-α*, *ura3Δ851*, *leu2Δ1*, *trp1Δ63*), in which *YKL222c* (position 3507–5621) or *YKR098c* (position 632659–634809) have been replaced by the *URA3* I-SceI cassette. Both strains are isogenic to the FY1679 (S288C) strain whose nuclear genome has been sequenced²⁶, except for the markers indicated. Both mutants retained the wild-type phenotype of the haploid cell. Yeast cells were transformed with the replicative I-SceI expression plasmid, pPEX7, by the lithium acetate method. Analysis of repair was performed by PCR directly on cells using 20-mer oligonucleotides whose 5'-end coordinates are 2725 and 6106 for FYBL2-5D *Δykl222c::URA3* I-SceI, and 632491 and 633481 for FYBL2-5D *Δykr098c::URA3* I-SceI (see Fig. 1). In both cases the primer distal to the centromere was biotinylated. Polymerase chain reaction products were separated on 1.5% agarose gels. Under these conditions insertions or deletions of at least 50 bp are detectable. These PCR-amplified DNA fragments were cut out from the gel and purified with QIAquick PCR Purification kit (Qiagen). The appropriate single-strand DNA was separated with streptavidine and then sequenced.

The *Δyme1* mutant of the strain FYBL2-5D *Δykr098c::URA3* I-SceI was obtained with the one-step gene-disruption technique by replacing *YPR024w* (*YME1*) with the KANMX module²⁹. The presence of the KANMX module instead of *YPR024w* was molecularly checked by PCR amplification of the specific fragment and by the restriction pattern of the amplified fragment with different restriction enzymes. As described for the *yme1* mutant¹⁹, the double mutant *Δyme1*, *Δykr098c* constructed here did not form ρ^- or ρ^0 colonies and was unable to grow on glucose at 14 °C and on glycerol at 37 °C. The ρ^- and ρ^0 mutants of the FYBL2-5D *Δykr098c* strain were obtained as described³⁰. By definition, ρ^- strains retain only a fraction of the wild-type mitochondrial genome, often in an amplified and rearranged configuration, and ρ^0 strains have no mitochondrial DNA. The ρ^- mutant was partially characterized by PCR amplification of short mitochondrial sequences and by Southern blot analysis using purified mitochondrial DNA digested with different restriction enzymes. These last two techniques were also used to reveal the absence of mitochondrial DNA in the ρ^0 mutants.

DNA sequence comparison

Sequences of PCR inserts were compared first with the Genbank and EMBL databases and, subsequently with the mitochondrial yeast genome⁴ using BLAST2. The systematic sequence comparison between the mitochondrial and the nuclear genome of FY1679 was performed using BLAST2. The nuclear yeast genome sequence was retrieved from MIPS (<http://speedy.mips.biochem.mpg.de/mips/yeast>) on 23 December 1998. The sequence of the mitochondrial genome of strain FY1679 was kindly provided by F. Foury⁴. Mitochondrial DNA sequence was fragmented into 200-bp-long sequences to serve as query. Two different fragmentations shifted by 100 bp with respect to each other were used. Defined mitochondrial sequences matching the nuclear genome were then compared once more with the nuclear genome and classified by increasing BLAST2 probability scores. Sequences with probability $\leq 6.0 \times 10^{-3}$ are shown in Table 2 and discussed in the text.

In addition to this first group of sequences, 137 other short mitochondrial sequences (18–21 bp in length) have 100% identity with nuclear sequences. They were excluded from the previous list based exclusively on BLAST scores but have statistical significance ($<10^{-3}$) if the length of the fragment, the size of the nuclear genome and the nucleotide composition of both the mitochondrial and the nuclear genome are taken into account. These sequences originate from both coding and non-coding regions of the mitochondrial genome. Although statistically significant, the fact that these sequences have such a small size and that a third of them are present in coding regions of chromosomes (compared with only 9% for the longer sequences) suggests that some of these may correspond to random coincidence rather than actual transfer. These sequences have not been taken into consideration in our conservative analysis.

Received 7 July; accepted 19 August 1999.

- Margulis, L. in *Origin of Eukaryotic Cells* (Yale University Press, New Haven and London, 1970).
- Perna, N. T. & Kocher, T. D. Molecular fossils in the nucleus. *Curr. Biol.* **6**, 128–129 (1996).
- Thorsness, P. E. & Weber, E. R. Escape and migration of nucleic acids between chloroplast, mitochondria, and the nucleus. *Int. Rev. Cytol.* **165**, 207–231 (1996).
- Foury, F., Roganti, T., Lecrenier, N. & Purnelle, B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**, 325–331 (1998).
- Resnick, M. A. & Martin, P. The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control. *Mol. Gen. Genet.* **143**, 119–129 (1976).
- Kramer, K. M., Brock, J. A., Bloom, K., Moor, J. K. & Haber, J. E. Two different types of double-strand breaks in *Saccharomyces cerevisiae* are repaired by similar RAD52-independent, nonhomologous recombination events. *Mol. Cell. Biol.* **14**, 1293–1301 (1994).
- Fairhead, C., Llorente, B., Denis, F., Soler, M. & Dujon, B. New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast* **12**, 1439–1457 (1996).
- Teng, S.-C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **383**, 641–644 (1996).
- Moore, K. J. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383**, 644–646 (1996).

- Fairhead, C., Thierry, A., Denis, F., Eck, M. & Dujon, B. 'Mass-murder' of ORFs from three regions of chromosome XI from *Saccharomyces cerevisiae*. *Gene* **223**, 33–46 (1998).
- Fairhead, C. & Dujon, B. Consequence of double-stranded breaks in yeast chromosomes: death or homozygosity. *Mol. Gen. Genet.* **240**, 170–180 (1993).
- de Zamaroczy, M. & Bernardi, G. The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae*—a review. *Gene* **47**, 155–157 (1986).
- Roth, D. B. & Wilson, J. H. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell. Biol.* **6**, 4295–4304 (1986).
- Roth, D. & Wilson, J. in *Genetic Recombination* (eds Kucherlapati, R. & Smith, G. R.) 621–653 (American Society for Microbiology, Washington DC, 1988).
- Gorbunova, V. & Levy, A. A. Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res.* **25**, 4650–4657 (1997).
- Sargent, R. G., Breneman, M. A. & Wilson, J. M. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* **17**, 267–277 (1997).
- Schiestl, R. H., Domiska, M. & Petes, T. D. Transformation of *Saccharomyces cerevisiae* with nonhomologous DNA: illegitimate integration of transforming DNA into yeast chromosomes and *in vivo* ligation of transforming DNA to mitochondrial sequences. *Mol. Cell. Biol.* **13**, 2697–2705 (1993).
- Thorsness, P. E. & Fox, T. D. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* **346**, 376–379 (1990).
- Thorsness, P. E. & Fox, T. D. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **134**, 21–28 (1993).
- Byers, B. in *The Molecular Biology of the Yeast Saccharomyces cerevisiae* (eds Strathern, J. N., Jones, E. W. & Broach, J. R.) 59–96 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1981).
- Farrelly, F. & Butow, R. Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301**, 296–301 (1983).
- Louis, E. J. & Haber, J. E. Evolutionarily recent transfer of a group I mitochondrial intron to telomere regions in *Saccharomyces cerevisiae*. *Curr. Genet.* **20**, 411–415 (1991).
- Blanchard, J. L. & Schmidt, G. W. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.* **13**, 537–548 (1996).
- Churcher, C. et al. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. *Nature* **387**, (Suppl) 84–87 (1997).
- Feuermann, M., De Montigny, J., Potier, S. & Souciet, J.-L. The characterization of two new clusters of duplicated genes suggests a 'lego' organization of the yeast *Saccharomyces cerevisiae* chromosomes. *Yeast* **13**, 861–869 (1997).
- Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
- Baudat, F. & Nicolas, A. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl Acad. Sci. USA* **94**, 5213–5218 (1997).
- Wolfe, K. M. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
- Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR based gene-disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
- Hauswirth, W. W., Lim, L. O., Dujon, B. & Turner, G. in *Mitochondria. A Practical Approach* (eds Darley-Usmar, V. M., Rickwood, D. & Wilson, M. T.) 171–242 (IRL Press, Oxford, 1987).

Acknowledgements

We thank F. Foury for having provided the entire mitochondrial sequence of *S. cerevisiae*, D. Alexandraki for the yeast strain *Δykl222c*, A. Harrington for purified mitochondrial DNA, A. Thierry for the *Δyme1* strategy, F. Tekia and A. Perrin for bioinformatics, M. Buckingham and T. Pugsley for comments on the manuscript and Henri Buc in whose laboratory some of these experiments were performed. This work was supported by a grant from the European Commission (EUROFAN). B.D. is a member of Institut Universitaire de France.

Correspondence and requests for materials should be addressed to M.R. (e-mail: mricch@pasteur.fr).

Mechanical unfolding intermediates in titin modules

Piotr E. Marszalek*, Hui Lu†, Hongbin Li*, Mariano Carrion-Vazquez*, Andres F. Oberhauser*, Klaus Schulten† & Julio M. Fernandez*

* Department of Physiology and Biophysics, Mayo Foundation, Rochester, Minnesota 55905, USA

† Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

The modular protein titin, which is responsible for the passive elasticity of muscle, is subjected to stretching forces. Previous work on the experimental elongation of single titin molecules has suggested that force causes consecutive unfolding of each domain in an all-or-none fashion^{1–6}. To avoid problems associated with the heterogeneity of the modular, naturally occurring titin, we

engineered single proteins to have multiple copies of single immunoglobulin domains of human cardiac titin⁷. Here we report the elongation of these molecules using the atomic force microscope. We find an abrupt extension of each domain by ~ 7 Å before the first unfolding event. This fast initial extension before a full unfolding event produces a reversible 'unfolding intermediate'. Steered molecular dynamics^{8,9} simulations show that the rupture of a pair of hydrogen bonds near the amino terminus of the protein domain causes an extension of about 6 Å, which is in good agreement with our observations. Disruption of these hydrogen bonds by site-directed mutagenesis eliminates the unfolding intermediate. The unfolding intermediate extends titin domains by $\sim 15\%$ of their slack length, and is therefore likely to be an important previously unrecognized component of titin elasticity.

The complex mechanical design of the elastic protein titin is still not understood^{1–7,10–20}. Recent *in vitro* studies using optical tweezers and atomic force microscopy (AFM) have revealed at least two components, where titin elasticity is governed by the entropic behaviour of its segments and the unravelling of its folded domains^{2,3,5}. The force–extension relationship of titin was explained by models of simple entropic elasticity^{2,3,5,7} in which module unfolding increased the contour length of the protein in a stepwise fashion^{2,5}. In the experiments using an atomic force microscope (AFM), extension of titin domains generated a force–extension relationship with a characteristic sawtooth pattern of equally spaced peaks that result from the sequential unfolding of its modules^{5,7}. In these recordings, the force–extension relationship of titin leading up to an unfolding event is described by the worm-like chain (WLC) model of polymer elasticity^{2,3,5–7}. However, close examination of the rising phases of the force peaks shows that they deviate from the shape predicted by the WLC model, raising the possibility that there are additional components in the elasticity of titin modules which have been missed. This is particularly evident in the rising phase of the first force peak which shows the elasticity of the protein before any unfolding event²¹.

Figure 1a shows that the force–extension curve of a native titin fragment strongly deviates from the expected entropic elasticity (WLC), revealing a pronounced 'hump' that tends to disappear on unfolding of all the modules. Native titin is a highly heterogeneous polymer, and therefore it is difficult to relate its elastic properties to individual structural elements. Hence, to examine the elastic components of titin modules at high resolution we engineered polyproteins composed of identical tandem repeats of either the I28 module (I28_n) or the I27 module (I27_n; ref. 7) of human cardiac titin (see Methods). Stretching a polyprotein with identical repeats of a module linearly amplifies the mechanical features of the protein module, revealing fine mechanical details at high resolution.

Force–extension curves for the I28_n and I27_n polyproteins showed the characteristic sawtooth patterns, corresponding to the sequential unfolding of their modules (Fig. 1b–d). Fits of the WLC to the force–extension curves leading up to the first peak (Fig. 1b–d; thin lines on the first peak) show a large deviation from simple entropic elasticity that becomes apparent at 108 ± 19 pN (I27, $n = 87$) and 151 ± 16 pN (I28, $n = 30$), well before the unfolding events occur (210 ± 27 pN, I27, $n = 87$; 264 ± 49 pN, I28, $n = 30$). When the force reaches ~ 100 – 150 pN (Fig. 1b–d), the slope of the force–extension curve decreases sharply creating the shape of a 'hump' that deviates from the line predicted by the WLC model. After this deviation, the slope of the curve increases again, after which the force reaches a maximum of about 200–300 pN where domain unfolding occurs.

Abrupt slope changes in the force–extension curve of a molecule, such as those shown in Fig. 1b–d, have been observed in polysaccharides and have been shown to correspond to force-induced transitions that elongate the molecule by a small amount^{6,22}. Hence, it is likely that the 'hump' observed in the force–extension curves of

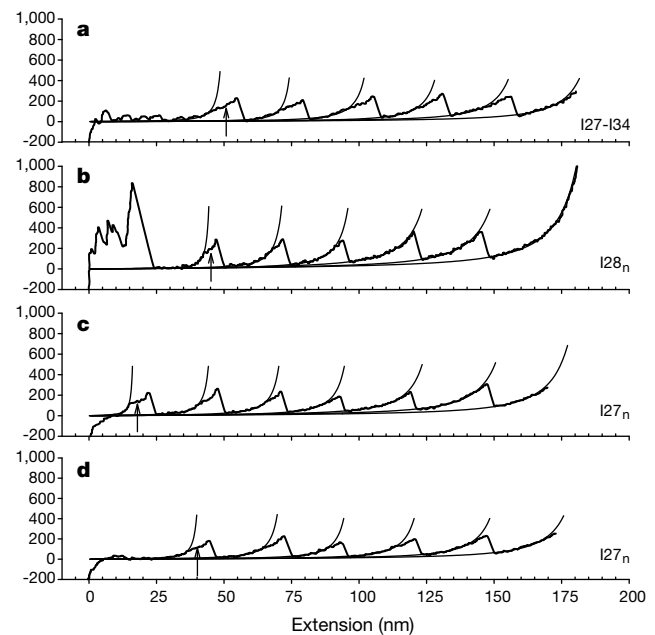


Figure 1 Force peaks corresponding to the sequential unfolding of immunoglobulin-like domains of human cardiac titin, showing large hump-like deviations from the WLC model of entropic elasticity (arrows). **a**, Force–extension curve for a fragment of human cardiac I band titin encompassing the immunoglobulin-like domains I27 to I34 (refs 5, 21).

b–d, Force–extension relationship for a polyprotein constructed from tandem repeats of the I28 module (**b**) and the I27 module (**c**, **d**). In all cases, the experimental data display a prominent hump in the rising phase of the initial force peaks and therefore can not be fitted with the WLC model (thin lines).

I28_n and I27_n corresponds to a previously undetected force-driven transition that elongates folded modules by a small amount at forces of 150 pN (I28) and 100 pN (I27). This elongation is a property of folded modules only. The hump in the force–extension relationship is most evident in the first peak, but it diminishes gradually with module unfolding and disappears completely when all the modules have unfolded and the protein is fully extended (Fig. 1; thin lines on the last peak).

The structure of the I27 module of titin is known, and its mechanical topology has been examined by molecular dynamics²³ and single-molecule AFM⁷. Thus, the I27 module is ideal to examine the molecular origin of the 'hump'. In Fig. 2, we analyse in detail the transitional extension observed at ~ 100 pN in the I27_n polyprotein. We measured the width of this transition as a contour length increment, ΔL_c , obtained by fitting the WLC model before and after the transition, as shown in Fig. 2a. These measurements were done on the transition observed in the first peak where the transition is most prominent. Although the I27_n polyprotein is engineered with 12 tandem modules, the AFM tip rarely picks up the protein from its termini. Hence, the AFM tip picks up proteins of various lengths, from very short proteins (2 modules long) up to their maximum length (12 modules long). We found that ΔL_c , measured in the first unfolding peak, increases linearly with the total number of folded modules of the polyprotein fragments picked by the AFM tip (Fig. 2b). A linear regression of the data in Fig. 2b gives a slope of 6.6 Å per module.

This result indicates that at ~ 100 pN, a folded domain undergoes a transitional extension of 6.6 Å; this extension contrasts with the much larger 284 Å extension caused by the full unfolding of the I27 module observed at ~ 200 pN (ref. 7). Figure 2c shows a histogram of the force at which the first transitional extension is observed compared with the unfolding force. A gaussian fit of the data (Fig. 2c; solid lines) gives a force of 108 pN for the small transitional extension, and 210 pN for the full unfolding event ($n = 87$).

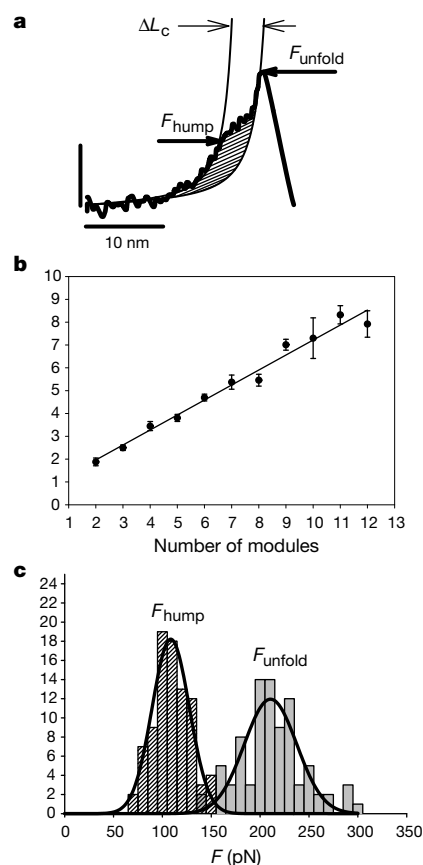


Figure 2 The size of the hump-like deviation depends linearly on the number of folded modules. **a**, Plot of the first force peak of a sawtooth pattern. The hump begins at a force, F_{hump} , that is smaller than the force required to completely unfold the module, F_{unfold} . The thin lines are fits of the WLC model to the data before and after the hump. The contour length of the second fit is $\Delta L_c = 46.5 \text{ \AA}$ longer than the one before the hump. The dashed area represents the work done by extending the protein through its intermediate. **b**, Relationship between ΔL_c (measured in the first unfolding peak) and the number of modules in the stretched tandem repeat. The solid line is a linear regression to the data with the slope: $6.6 \text{ \AA}$ per module ($r^2 = 0.98$, $n = 87$). The bars indicate standard errors. **c**, Histogram of F_{hump} and F_{unfold} . Gaussian fits (solid lines) to the data give average values of $F_{\text{hump}} = 108 \text{ pN}$ and $F_{\text{unfold}} = 210 \text{ pN}$ ($n = 87$). The pulling speed was $0.3\text{--}0.5 \text{ nm ms}^{-1}$.

However, the unfolding forces are dependent on the rate at which the protein is pulled^{5–7,24}. Hence it is likely that the small transitional extension observed here may also occur at much lower forces.

Although the small transitional extension of the I27 module is most prominent in the first unfolding peak of the force–extension curve, this transition is also observed in consecutive peaks, albeit with decreasing amplitude (Fig. 1). These observations put together suggest that the transition observed in the first peak corresponds to the accumulated transitional extension of all the folded domains. The reappearance of a progressively smaller transition in consecutive peaks suggests that the transition is reversible upon relaxation of the force (peak to trough transition), and that a small transitional extension of the modules that remain folded occurs again at a force of $\sim 100 \text{ pN}$. We estimate that the refolding rate of the small transitional extension must be at least 25 s^{-1} in order to fully recover between unfolding events (Fig. 1). This fast refolding rate contrasts with the much slower refolding observed after full unfolding of the I27 module ($\sim 1 \text{ s}^{-1}$; ref. 7).

Steered molecular dynamics (SMD) simulations of the forced extension of the I27 module predict that the mechanical resistance in this module arises from a patch of six hydrogen bonds bridging the A' and G β -strands (Fig. 3a). According to these simulations, the

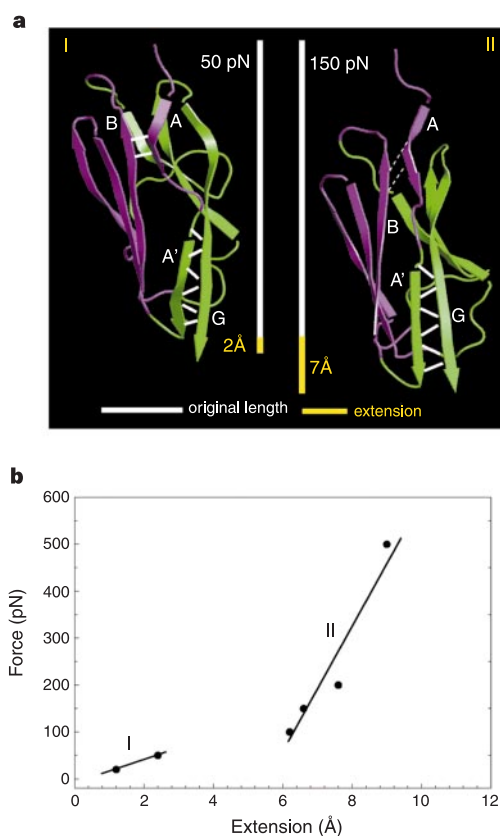


Figure 3 Steered molecular dynamics simulations of I27 extensibility under constant force. **a**, ‘Snapshots’ of the structure of the I27 module simulated at a force of 50 pN (I, at 1 ns) and 150 pN (II, at 1 ns). At 50 pN the hydrogen bonds between strands A and B are maintained, whereas at 150 pN they are broken. **b**, I27 force–extension relationship obtained from the simulations. The discontinuity observed between 50 and 100 pN corresponds to an abrupt extension of the module by $4\text{--}7 \text{ \AA}$ caused by the rupture of the AB hydrogen bonds, and the subsequent extension of the partially freed polypeptide segment. Figure generated using the program VMD³⁰.

simultaneous rupture of this patch of hydrogen bonds is the critical event that allows the full unravelling of the module under an applied force²³. These simulations showed that the rupture of the two hydrogen bonds bridging the A and B β -strands was a minor event that preceded the rupture of the A'G patch²³. We examined whether the small extension that precedes the main unfolding event was due to the rupture of two hydrogen bonds bridging the AB β -strands (K6–E24). For this purpose, SMD simulations were performed in which forces of 20, 50, 100, 150, 200 and 500 pN were applied, each for 1 ns, to the termini of a single I27 domain. Upon application of these forces, the I27 domain extended within less than 200 ps to a quasi-equilibrium value that is plotted in Fig. 3b. Full unfolding within the 1-ns time window was observed only for the strongest force (500 pN). At forces of less than 50 pN the I27 module was forced to extend by $\sim 2 \text{ \AA}$, whereas for forces larger than 100 pN the quasi-stationary extensions ranged from 6 to $9 \text{ \AA}$. Analysis of the SMD trajectories showed that at forces of 20 pN and 50 pN the hydrogen bonds between the A and B β -strands are maintained (Fig. 3b, line I) and that the $2\text{-\AA}$ extension arising in this case is due to straightening of the terminal residues. At larger forces these hydrogen bonds break (Fig. 3b, line II), producing a discontinuous extension from $2 \text{ \AA}$ at lower forces to an additional $4\text{--}7 \text{ \AA}$ (at forces $\geq 100 \text{ pN}$; Fig. 3b). The discontinuous elongation caused by the rupture of these hydrogen bonds is in good agreement with the $6.6\text{-\AA}$ elongation observed with the AFM. This event occurs at low force and precedes the full unravelling of the I27 module. Furthermore, an estimate of the mechanical work done to extend

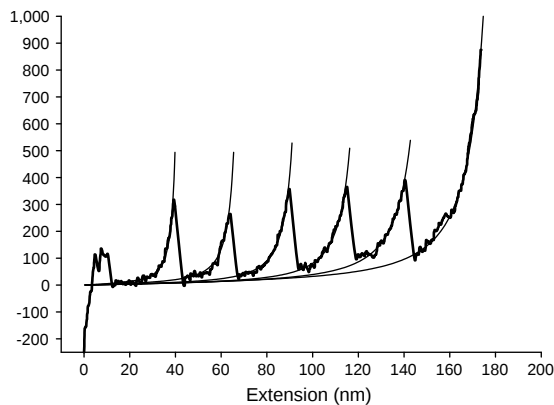


Figure 4 Removal of the deviation from the WLC model by introducing a point mutation in the I27 polyprotein (I27₉-K6P). The point mutation substituted lysine 6 by a proline. This proline in position 6 disrupts the two critical hydrogen bonds that link the A and B β -strands together. In contrast to the wild-type form of the I27 module, the rising phase of all force peaks measured from the I27₉-K6P protein closely follows the WLC model.

the modules at low force gives a value of 10.1 ± 0.63 kcal per mol per module (Fig. 2a, hatched area) which agrees well with the work required to rupture a pair of hydrogen bonds²⁵.

These observations strongly suggest that the unfolding intermediate that we observe during the extension of the I27 module corresponds to the rupture of the pair of hydrogen bonds bridging the A and B β -strands of the I27 module. To verify this view, we engineered a point mutant polyprotein (I27₉-K6P; 9 repeats of a mutant I27 module where lysine in position 6 was replaced by a proline) where the mutation interferes with the formation of these hydrogen bonds. Figure 4 shows a force-extension curve for the (I27₉-K6P) polyprotein. The figure shows that, in contrast to the wild-type form, the mutant polyprotein does not show deviations from the WLC model. Hence, introduction of the K6P mutation eliminates the unfolding intermediate that we have described.

A folded I27 module has a length of 44 Å, as determined by NMR¹³. We have shown that a force of ~ 100 pN can extend the module by 6.6 Å without causing full domain unfolding. This rapidly reversible unfolding intermediate corresponds to an extension of a titin module by 15% of its resting length. Our data therefore provide direct evidence for a new source of extensibility in human cardiac titin, caused by a force-driven conformational change in its immunoglobulin domains. □

Methods

Protein engineering

We constructed polyproteins containing 12 direct tandem repeats of a single immunoglobulin domain no. 27, and 8 tandem repeats of a single immunoglobulin domain no. 28 from the I band of human cardiac titin, according to methods described elsewhere⁷. We also expressed and purified the I27-134 construct of human cardiac titin synthesized elsewhere⁵. The I27-K6P point mutation monomer was generated by polymerase chain reaction using mutagenic primers. All these recombinant proteins have two Cys residues in their carboxy terminus to facilitate the attachment of the molecules to the gold-coated coverslips.

Force spectroscopy of single proteins

Our custom-made AFM apparatus and its mode of operation have been described²⁶. Calibration of the spring constant of each individual cantilever was done in solution using the equipartition theorem as described in ref. 27. The proteins were suspended in PBS buffer at a concentration of $\sim 10 \mu\text{g ml}^{-1}$ and adsorbed onto freshly evaporated gold coverslips. For the analysis of the polyproteins, we chose only recordings showing a sawtooth force pattern that ended with a high-force peak that corresponds to stretching a fully denatured protein (all modules unfolded). This allowed us to count the number of modules within the fragment of the protein that were actually stretched.

Steered molecular dynamics simulation

The steered molecular dynamics (SMD) simulations of I27 were performed with the programs NAMD²⁸ and XPLOR²⁹ with the CHARMM22 (ref. 29) force field. The structure

of I27 has been reported¹³ and is available elsewhere (Brookhaven PDB databank; PDB code 1TIT). The structure was solvated in water and equilibrated following the procedure described elsewhere²³. SMD simulations with constant forces were performed by fixing one terminus of the domain and applying the force to the other terminus. The direction of the force was chosen along the vector from the N terminus to the C terminus. The constant stretching force F (pN) was determined by adding a harmonic potential between the C terminus and a point relatively far away (400 nm) from it along the force direction with the spring constant set to $F(\text{pN})/400(\text{nm})$. The fluctuation of the magnitude of the force is less than 0.2%. The simulation adopted a timestep of 1 fs, a uniform dielectric constant of 1, and a cut-off of Coulomb and van der Waals interactions with switching function starting at a distance of 10 Å and reaching zero at 13 Å. Simulation time, 1 ns.

Received 26 April; accepted 18 August 1999.

- Erickson, H. P. Stretching single protein molecules: titin is a weird spring. *Science* **276**, 1090–1092 (1997).
- Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
- Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L. & Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116 (1997).
- Kellermayer, M. S. Z., Smith, S. B., Bustamante, C. & Granzier, H. L. Complete unfolding of the titin molecule under external force. *J. Struct. Biol.* **122**, 197–205 (1998).
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
- Rief, M., Fernandez, J. M. & Gaub, H. E. Elastically coupled two-level-systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **81**, 4764–4767 (1998).
- Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
- Lu, H. & Schulten, K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins Struct. Funct., Genet.* **35**, 453–463 (1999).
- Izrailev, I., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72**, 1568–1581 (1997).
- Soteriou, A., Clarke, A., Martin, S. & Trinick, J. Titin folding energy and elasticity. *Proc. R. Soc. Lond. B* **254**, 83–86 (1993).
- Politou, A. S., Thomas, D. J. & Pastore, A. The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys. J.* **69**, 2601–2610 (1995).
- Politou, A. S., Gautel, M., Improta, S., Vangelista, L. & Pastore, A. The elastic I-band region of titin is assembled in a “modular” fashion by weakly interacting Ig-like domains. *J. Mol. Biol.* **255**, 604–616 (1996).
- Improta, S. *et al.* The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. *J. Mol. Biol.* **284**, 761–777 (1998).
- Granzier, H., Helmes, M. & Trombitas, K. Nonuniform elasticity of titin in cardiac myocytes: A study using immunoelectron microscopy and cellular mechanics. *Biophys. J.* **70**, 430–442 (1996).
- Granzier, H., Kellermayer, M., Helmes, M. & Trombitas, K. Titin elasticity and mechanism of passive force development in rat cardiac myocytes probed by thin-filament extraction. *Biophys. J.* **73**, 2043–2053 (1997).
- Trombitas, K. *et al.* Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. *J. Cell Biol.* **140**, 853–859 (1998).
- Labeit, S. & Kolmerer, B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
- Labeit, S., Kolmerer, B. & Linke, W. A. The giant protein titin. Emerging roles in physiology and pathophysiology. *Circ. Res.* **80**, 290–294 (1997).
- Linke, W. A., Stockmeier, M. R., Ivemeyer, M., Hosser, H. & Mundel, P. Characterizing titin's I-band Ig domain region as an entropic spring. *J. Cell Sci.* **111**, 1567–1574 (1998).
- Gautel, M. & Goulding, D. A molecular map of titin/connectin elasticity reveals two different mechanisms acting in series. *FEBS Lett.* **385**, 11–14 (1996).
- Rief, M., Pascual, J., Saraste, M. & Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.* **286**, 553–561 (1999).
- Marszalek, P. E., Oberhauser, A. F., Pang, Y.-P. & Fernandez, J. M. Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature* **396**, 661–664 (1998).
- Lu, H., Israilewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671 (1998).
- Evans, E. & Ritchie, K. Strength of a weak bond connecting flexible polymer chains. *Biophys. J.* **76**, 2439–2447 (1999).
- Hermans, J., Berendsen, H. J. C., Van Gunsteren, W. F. & Postma, J. P. M. A consistent empirical potential for water-protein interactions. *Biopolymers* **23**, 1513–1518 (1984).
- Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).
- Florin, E. L. *et al.* Sensing specific molecular interactions with the atomic force microscope. *Biosens. Bioelectr.* **10**, 895–901 (1995).
- Nelson, M. *et al.* NAMD—A parallel, object-oriented molecular dynamics program. *Int. J. Supercomputer Applications and High Performance Computing* **10**, 251–268 (1996).
- Brunger, A. X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR (Yale Univ., New Haven & London, 1992).
- Humphrey, W. F., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

Acknowledgements

We thank M. Gautel for a gift of the I27-134/pET8c plasmid, and T. Fisher for comments on the manuscript. This work was supported by NIH and NSF grants to J.M.F., P.E.M. A.F.O. and K.S., and by the Roy J. Carver Charitable Trust to K.S.

Correspondence and requests for materials should be addressed to J.M.F. (e-mail: fernandez-julio@mayo.edu).

Molecular matters

ATP, DNA, SPC and other TLAs (three-letter acronyms) feature this week.



Leave it: the GENESIS workstation can go it alone.

GENESIS workstation

Tecan

www.tecan.ch

A workstation for the molecular biologist

The new GENESIS Molecular Biology Workstation is a 'walk-away' system suitable for automating DNA extraction and sample preparation for PCR or sequencing. It can be adapted to use both solid-phase extraction and magnetic-bead separation methods. The liquid-handling arm can pipette a wide range of substances, such as lysing solutions, magnetic beads and samples, into tubes or 96-well microplates. A low-volume option ensures reliable pipetting of volumes as low as 0.5 µl.

Reader Service No. 100

ElectroZap

Ambion

www.ambion.com

Eliminate RNase and DNase contamination from glass and plastic surfaces

ElectroZap is a noncorrosive solution designed for the decontamination of pH meter electrodes. Electrodes need to be RNase-free for the preparation of solutions that cannot be DEPC-treated, such as Tris-HCl. ElectroZap is simply sprayed onto the electrode and then rinsed off with RNase-free water.

Reader Service No. 101

PARP

Trevigen

www.trevigen.com

Two takes on PARP

Two grades of PARP (poly(ADP-ribose) polymerase) are available from Trevigen: a 'standard' PARP, and a new high-specific-activity enzyme. The new protein has a specific activity of greater than 5×10^5 U per mg. Standard PARP enzyme is useful as a positive control for both western blot analysis and the PARP activity assay. For more detailed studies, the purified PARP is recommended.

Reader Service No. 102

GeneSpec I

Web Scientific

www.webscientific.co.uk

DNA/RNA quantitation in just 1 µl

Volumes down to 1 µl, and results in 8 seconds are claimed for the GeneSpec I mini-spectrophotometer. This diode-array-based machine is capable of providing A_{260}/A_{280} , DNA/ssDNA/dsDNA concentration, purity, molar concentration and protein concentration instantly. Data and spectra may be stored and displayed via a PC. GeneSpec I occupies a space the size of an A4 sheet in bench area.

Reader Service No. 103

ZyMAX

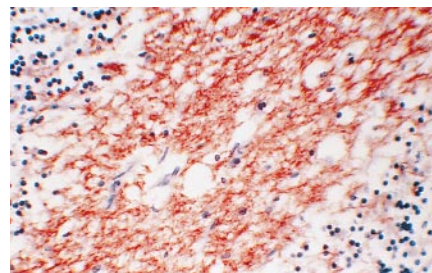
Zymed Laboratories

www.zymed.com

'Second generation' streptavidin conjugates

These new streptavidin-horseradish peroxidase and streptavidin-alkaline phosphatase conjugates have at least a 2-fold higher titre than standard grade. Streptavidin has replaced avidin as the reagent of choice for most applications where protein interactions may cause background. ZyMAX conjugates are recommended for immunochemistry, western blotting, ELISA, flow cytometry, immunofluorescence and other immunochemical techniques.

Reader Service No. 104



Back to basics: MBP in the cerebellum.

Myelin antibody

InnoGenex

www.innogenex.com

Antibody to a major component of myelin

Monoclonal 10F4 (IgG1/κ) was produced using a synthetic peptide (ASQKRPSQRH-GSK) encompassing amino acids 2-14 near the N terminus of human myelin basic protein (MBP). The sequence of the immunizing peptide has 100% identity with rat, chimpanzee, pig and guinea pig MBP, predicting cross-reactivity with these species. 10F4 is recommended for immunostaining, ELISA and western blotting.

Reader Service No. 105

Cytokines poster

Serotec

www.serotec.co.uk

Free, while stocks last

The 'A-to-Z of Cytokines' poster is now available from Serotec. The poster gives a reference guide to the classic cytokines, growth factors, chemokines and their receptors. Recent introductions to the Serotec range include apolipoprotein A, eotaxin and exodus 2, FLT-3 ligand, lymphotactin and oncostatin M.

Reader Service No. 106

Ageing antibodies

Autogen Bioclear

www.autogen-bioclear.co.uk

Antibodies for the study of ageing and molecular clock/circadian rhythms

This new antibody range includes antibodies

ADVERTISEMENT

READER ENQUIRY NO. 26

to Klotho, telomerase and associated proteins (Est2, TRF1, TRF2, TP1), Clock, mouse and human Per 1, 2 and 3, BMAL, MOP3 and 4, and cryptochromes (CRY1 and CRY2).

Reader Service No. 107

Anti-6HIS

Packard Instrument Company

www.packardinst.com

Two epitope-tag recognition reagents for assay development and high-throughput screening

Anti-6HIS-Eu(K) is an IgG2a mouse monoclonal antibody to the 6HIS tag conjugated with europium cryptate, the primary fluorophor of HTRF (homogeneous time-resolved fluorescence) chemistry. Anti-6HIS-XL665 is an anti-6HIS antibody conjugated with the secondary fluorophor of HTRF, cross-linked allophycocyanin. These reagents are suitable for high-throughput screening of protein/protein interactions and receptor/ligand binding assays.

Reader Service No. 108

ATP assay

LabSystems

www.labsystems.com

Make light work

The Bioluminescent ATP Assay Kit from LabSystems Affinity Sensors is based on the luciferase-catalysed reaction, where light emitted is directly proportional to ATP concentration. The kit contains all reagents and disposable materials required and provides a linear, time-independent light output which can be measured using a luminometer. The assay is extremely sensitive, with no amplification step needed to bring cell numbers within the scale of sensitivity. Intracellular ATP can easily be assayed using the ATP-releasing agent included in the kit, the detergent-based reagent being suitable for a wide variety of bacterial and mammalian cells.

Reader Service No. 109

Reagents in a box

Eppendorf

www.eppendorf.de

Ready-to-use reagents for molecular biology

Electrophoresis and blotting reagents from Eppendorf can now be obtained in 4- and 10-



Box clever: stackable reagents from Eppendorf.

litre dispenser boxes. The boxes, not unlike those favoured by oenologists, are fitted with dispenser valves, and the reagents inside are guaranteed to be sterile and free of RNase and DNase.

Reader Service No. 110

Spectra Online

Galactic Industries

www.galactic.com

Add '/spconline' to the URL for searchable spectra

Software developer Galactic Industries has posted a free searchable database on the company's website. Visitors to Spectra Online can query the database by submitting spectra in Galactic SPC format for a full point-by-point spectra match. Parameters that can be searched include compound name, CAS number, molecular formula, molecular weight, melting point and boiling point.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

CUSTOM PEPTIDES

• Up to 30 aa
• HPLC & MS data
• Purity >80% >95% >99%
• Amount 15mg 10mg

info@genosphere-biotech.com
Tel: +33 1 42 71 70 21 • Fax: +33 1 42 71 79 23

GENOSPHERE BIOTECHNOLOGIES

READER ENQUIRY NO. 24

contact: lightcycler@TIB-MOLBIOL.de

TIB MOLBIOL

HybProbes
for the LightCycler

■ Mutation Detection
■ Quantitative PCR

Custom Synthesis of Hybridization Probes for the new LightCycler under license of Roche Diagnostics

READER ENQUIRY NO. 7

Silver Stain Kit II
for Electrophoresis
Rapid, sensitive detection of proteins & nucleic acids

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1000 Belwood Road
Fremont, CA 94538
Telephone: (925) 271-7077
Fax: (925) 271-7701
(800) 922-6226
wako@wako.com

READER ENQUIRY NO. 41

luminometer.committed™

luminometer.com

A website of luminometers and their application methods.

TURNER DESIGNS
845 W. Maude Avenue • Sunnyvale, CA 94386
(408) 745-0964 • FAX: (408) 745-0968

READER ENQUIRY NO. 10

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING
Microverse-based, solid or solution samples
£12 per cycle
5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray
Phone/Fax: +44 191 222 8512
Email: JOE.GRAY@newcastle.ac.uk

DNA sequencing (AB1 577-8882, 577 Colony) also available at competitive prices

READER ENQUIRY NO. 2

Check out our homepage at <http://www.resgen.com>

**BACs
YACs
PACs
cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 6

Homogenizers?

■ Tissue-Tearor™
■ BeadBeater™
■ Mini-BeadBeater™
■ BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail: info@biospec.com

New Probe Sizes
0.45cm!

READER ENQUIRY NO. 1